# (DRAFT OF) APPLICATIONS IN LINEAR ALGEBRA FOR MATH 223, UBC, SPRING 2019

JOEL FRIEDMAN

ABSTRACT. This article gives a number of applications of linear algebra without assuming any knowledge of linera algebra (beyond solving a system of two linear equations of two unknowns, typically done in high school). It will be used in the first few weeks of Math 223 to motivate linear algebra; it will also be used throughout the course as a source of examples.

## CONTENTS



*Date*: Tuesday 22nd January, 2019, at 09:40(get rid of time in final version).
2010 *Mathematics Subject Classification.* Primary .
Research supported in part by an NSERC grant.

THIS ARTICLE IS CURRENTLY IN DRAFT FORM. SOME MISTAKES IN THIS ARTICLE WILL ONLY BE CORRECTED IN CLASS.

## 0. Introduction

For the first few weeks of Math 223 course (Spring 2019) we describe some examples and applications of linear algebra.

We are inspired by Prof. Klaus Hoechsmann, who began his linear algebra courses with a two-week introduction covering the entire course content in the special case of $2 \times 2$ matrices; this way the students would understand part of the "big picture" of linear algebra before starting with the technicalities. Our particular fascination with linear algebra is its diversity of applications, so we begin will begin Math 223 with some representatative applications; this is another way to get the "big picture."

In addition, Prof. Kai Behrend suggested that we provide more applications to supplement the current Math 223 textbook.

These notes do not assume any knowledge of linear algebra beyond the ability to solve $2 \times 2$ linear systems. Terms in italics, such as *kernel* or *quadratic form*, are terms that we will eventually define precisely and study theoretically in Math 223; in this article we don't require the reader to understand these terms. Some of the exercises in these notes ask the reader to type some square matrices (arrays of numbers) into linear algebra software and take powers of the matrices; we don't assume that the reader knows how to multiply matrices, rather we explain what these powers mean in practical terms.

At the end of a phrase or sentnence, we use the term "(exercise)" to mean that we intend to create an exercise (with precise statements and possibly some hints) based on what is written; we use the term "(why?)" to suggest that the reader might be able to figure out why we claim is true. Some terminology we use, such as *linearity* or *nice* (e.g., sufficiently *nice* function) are purposely vague; much of this vagueness will be clarified later in Math 223.

These notes have a lot of applications, too many to cover in two or three weeks; some of these applications will be covered later in the course, to illustrate concepts in linear algebra as we cover them.

**At this point these notes are a work in progress; exercises and material may be added throughout the course.** Part of these notes are rather skeletal, and some details may only be given in class. In particular, **many helpful diagrams and examples may only be given in class**. Also, **some errors in these notes may only be fixed in class**; anyone not taking Math 223 is welcome to read these notes (but should be aware of possible errors and incomplete explanations).

Remarks on the current state of these notes:

(1) At present a number of sections or subsections have no material.
(2) Some exercises are interspersed in the main text; other additional exercises appear at the end of the sections. Most likely I will add more exercises, and **some exercise numbers may change** until I assign homework from that section.

We finish this section with some preliminary notation and background.

0.1. **Notation.** We use $\mathbb{R}$ to denote the real numbers, and

$$\mathbb{N} = \{1, 2, 3, \ldots\}$$

to denote the natural numbers, and

$$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$$

to denote the integers, and the notation

$$\mathbb{Z}_{\geq 0} = \{0, 1, 2, \ldots\} .$$

**Warning:** Many computer science programming languages use $\mathbb{Z}_{\geq 0}$ for the natural numbers, i.e., count starting from 0 (why?); many math and computer science texts also use this convention.

0.2. **The Complex Numbers.** There are many good introductions to the complex numbers; the current Wikipedia does a good job, but there are more succint expositions.

We use $i$ to denote $\sqrt{-1}$; the complex numbers are formal combinations

$$\mathbb{C} = \{\ a + bi \mid a, b \in \mathbb{R}\}.$$

I assume you know how to add, subtract, multiply, and divide complex numbers, e.g.,

$$\begin{aligned}\frac{a+bi}{c+di} &= (a+bi)\frac{1}{c+di} = (a+bi)\frac{1\cdot(c-di)}{(c+di)\cdot(c-di)} = (a+bi)\frac{c-di}{c^2+d^2} \\ &= \left(\frac{ac-bd}{c^2+d^2}\right) + \left(\frac{bc-ad}{c^2+d^2}\right) i\ .\end{aligned}$$

Later in Math 223 we will use the fact that any nonzero complex number $a+bi$ has a unique (*polar*) representation as $re^{i\theta}$ for $r > 0$ and $\theta \in [0, 2\pi)$, i.e., $0 \le \theta < 2\pi$; if

$$a + bi = re^{i\theta} \quad \text{then} \quad (a+bi)^n = r^n e^{i\theta n}$$

for any $n \in \mathbb{Z}$. This is useful to know when $a, b$ are fixed and $n \to \infty$ (or $n \to -\infty$).

0.3. **Summation Notation.** If $f(i)$ is a function of $i \in \mathbb{Z}$ (with values in $\mathbb{C}$, $\mathbb{Z}$, or any values where there is a reasonable notion of addition), and $a \le b$ are integers, we use the notation

$$\sum_{i=a}^{b} f(i) \quad \text{to denote} \quad f(a) + f(a+1) + f(a+2) + \cdots + f(b);$$

in particular

$$\sum_{i=a}^{a} f(i) = f(a).$$

At times it is useful to define the above summation when $b < a$; by convention,

$$\sum_{i=a}^{a-1} f(i) = 0, \quad \sum_{i=a}^{a-2} f(i) = -f(a-1), \quad \sum_{i=a}^{a-3} f(i) = -f(a-1) - f(a-2), \quad \cdots$$

(why is this a reasonable definition?). The abstract idea allows us to extend many common sequences of numbers backwards; this is useful in Subsection 3.1.

0.4. **Product Notation.** Similarly we write

$$\prod_{i=a}^{b} f(i) = f(a)f(a+1)\cdots f(b)$$

with

$$\prod_{i=a}^{a-1} f(i) = 1, \quad \prod_{i=a}^{a-2} f(i) = \frac{1}{f(a-1)}, \quad \prod_{i=a}^{a-3} f(i) = \frac{1}{f(a-1)f(a-2)}, \quad \cdots$$

In these notes we prefer to avoid these summation and product notation whenever reasonably possible, but this notation is at times both conceptually and notationally better to use.

0.5. **Proofs By Induction.** A lot of facts we will use in our examples are theorems that can easily be proven by induction.

Formally, induction works via the following principle: consider a set $S \subset \mathbb{N}$ for which (1) $1 \in S$, and (2) $i \in S$ implies $i+1 \in S$. Then $S = \mathbb{N}$.

**Example 0.1.** Let us prove that for all $n \in \mathbb{N}$,

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}, \tag{1}$$

i.e.,

$$1 + 4 + 9 + 16 + 25 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6} \ . \tag{2}$$

Let $S$ be the set of $n \in \mathbb{N}$ for which the above formula is true. Then (2) holds for $n = 1$ (since $1 = 1(1+1)(2+1)/6$); hence $1 \in S$. Assuming that $i \in S$, i.e., that (2) holds for $n = i$, we have

$$1 + 4 + 9 + \cdots + i^2 = \frac{i(i+1)(2i+1)}{6};$$

adding $(i+1)^2$ to both sides we have

$$1 + 4 + \cdots + i^2 + (i+1)^2 = \frac{i(i+1)(2i+1)}{6} + (i+1)^2 = \frac{(i+1)(i+2)(2i+3)}{6}$$

which is just (2) for $n = i+1$. Hence $i \in S$ implies that $i+1 \in S$. It follows that $S$, the set of $n$ for which (2) holds, is all of $\mathbb{N}$.

These types of "proof by induction" are often written informally, as proving a theorem by proving the *base case* (i.e., $1 \in S$), and the *inductive hypothesis* ($i \in S \Rightarrow i+1 \in S$). In Math 223, we will insist that proofs by induction are written in the formal manner of Example 0.1.

0.6. **(Additional) Exercises.**

I am in the procdss of adding more exercises to this part and/or changing their order; THE EXERCISE NUMBERS MAY CHANGE until I assign homework from this part.

N.B. All proofs by induction must be written as follows: (1) they must begin with "Let $S \subset \mathbb{N}$ be the set of $n$ such that ..." (2) must continue with "Let us first show that $1 \in S$: ..." (3) must continue with "Let us next show that $i \in S$ implies that $i+1 \in S$: ..." (4) must end with "Since $1 \in S$ and $i \in S \Rightarrow i+1 \in S$, it follows that $S = \mathbb{N}$; it follows that ..." NO CREDIT WILL BE GIVEN TO PROOFS THAT DO NOT FOLLOW THIS FORMAT.

**Exercise 0.1.** For $n \in \mathbb{N}$, with $n \geq 2$, $n$ curling teams play one another in a single elimination tournament. (In the first match some team plays some other; the loser of the match is eliminated from the tournament; the winner survives and continues in the tournament with the other $n-2$ players.) Use induction to show that for all $n \in \mathbb{N}$, a single elimination tournament with $n+1$ terms will end with a single winning team after $n$ matches (no matter the order that the teams play one another). [It is easy to prove this fact directly; the point of this exercise is to get practice formally writing out proofs by induction. Make sure you use the formal form of induction above.]

**Exercise 0.2.** Use induction to show that for all $n \in \mathbb{N}$

$$1 + 3 + 5 + \cdots + (2n-1) = n^2 \tag{3}$$

[It is easy to verify (3) by different means; the point of this exercise is to get practice formally writing out proofs by induction. Make sure you use the formal form of induction above.]

**Exercise 0.3.** Use induction to show that for all $n \in \mathbb{N}$

$$1 + 2^3 + 3^3 + \cdots + n^3 = \frac{n^2(n-1)^2}{4} \tag{4}$$

[It is easy to verify (4) by different means; the point of this exercise is to get practice formally writing out proofs by induction. Make sure you use the formal form of induction above.]

**Exercise 0.4.** Fix $x \in \mathbb{R}$. Show that for all $n \in \mathbb{N}$ we have

$$(1-x)\big(1 + x + \cdots + x^n\big) = 1 - x^{n+1} \tag{5}$$

[It is easy to directly verify (5); the point of this exercise is to get practice formally writing out proofs by induction. Make sure you use the formal form of induction above.]

**Exercise 0.5.**

0.5(a) Fix $\theta \in \mathbb{R}$. Use induction to show that for $n \in \mathbb{N}$ we have

$$2\sin(\theta)\big(\cos(0) + \cos(2\theta) + \cos(4\theta) + \cdots + \cos(2n\theta)\big) = \sin\big((2n+1)\theta\big) \tag{6}$$

[Hint: You don't have to know any trig; just use the identity

$$\sin(\alpha + \theta) - \sin(\alpha - \theta) = 2\cos(\alpha)\sin(\theta)$$

for appropriately chosen $\alpha$.] [It is easy to directly verify (6); the point of this exercise is to get practice formally writing out proofs by induction. Make sure you use the formal form of induction above.]

0.5(b) Recall that

$$\lim_{\theta \to 0} \frac{\sin\theta}{\theta} = 1$$

(when sin is measured in radians). Show by taking Riemann sums that for any $\eta \in \mathbb{R}$, $\int_0^\eta \cos(x)\,dx = \sin(\eta)$ (when measuring angles radians).

**Exercise 0.6.** Let $F_n$ denote the $n$-th Fibonacci number, i.e., $F_1 = 1$, $F_2 = 1$, and for $n \geq 3$ we have $F_n = F_{n-1} + F_{n-2}$. Hence

$$F_1 = 1,\ F_2 = 1,\ F_3 = 2,\ F_4 = 3,\ F_5 = 5,\ F_6 = 8,\ F_7 = 13,\ F_8 = 21, \ldots$$

(some texts will enumerate the Fibonacci numbers in slightly differing notation and indexing).

0.6(a) Apply the formula $F_n = F_{n+2} - F_{n+1}$ for $n = 0, -1, -2, \ldots, -8$ to derive values for $F_0, F_{-1}, \ldots, F_{-8}$; *guess* a simple formula (i.e., write down a simple formula for the pattern you see) for $F_n$ when $n$ is a negative integer, and prove by induction that your guess is correct.

0.6(b) Compute $F_n F_{n+2} - F_{n+1}^2$ for $n = 1, \ldots, 5$; guess a simple formula (i.e., write down the pattern you see) for this expression, and prove by induction that your guess is correct.

0.6(c) Compute $F_nF_{n+3} - F_{n+1}F_{n+2}$ for $n = 1, \dots, 5$; guess a simple formula (i.e., write down the pattern you see) for this expression, and prove by induction that your guess is correct.

0.6(d) Compute $F_nF_{n+8} - F_{n+1}F_{n+7}$ for $n = 1, \dots, 5$; you can write your answer in terms of the Fibonacci numbers [by substituting $F_{n+7} + F_{n+6}$ for $F_{n+8}$ and $F_n + F_{n-1}$ for $F_{n+1}$] rather than writing out the actual integer; guess a simple formula (i.e., write down the pattern you see) for this expression, and prove that your guess is correct.

0.6(e) Compute $F_nF_{n+100} - F_{n+1}F_{n+99}$ for $n = 1, \dots, 5$; you can write your answer in terms of the Fibonacci numbers [see the previous part] rather than writing out the actual integer; guess a simple formula (i.e., write down the pattern you see) for this expression, and prove that your guess is correct.

0.6(f) Prove by induction that if $\xi_+ = (1 + \sqrt{5})/2$ and $\xi_- = (1 - \sqrt{5})/2$, then $F_n = (\xi_+^n - \xi_-^n)/\sqrt{5}$. [Hint: first find the roots of $x^2 = x + 1$.]

0.6(g) Use the above formula to prove that for $n \geq 0$, $F_n$ is the integer nearest to $\xi_+^n/\sqrt{5}$.

0.6(h) Prove that the GCD (greatest common divisor) of $F_n$ and $F_{n+1}$ is 1.

**Exercise 0.7.** Recall that the number of subsets of size $k$ from a fixed set of $n$ elements is

$$\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k(k-1)\dots 1} = \frac{n!}{k!(n-k)!},$$

where ! denotes the "factorial," e.g., $k! = k(k-1)\dots 1$.

0.7(a) Prove that for any $n \in \mathbb{N}$ we have that $\sum_{m=1}^{n} m = 1 + 2 + \dots + n$ equals $\binom{n+1}{2}$; use induction on $n$.

0.7(b) Prove that for any $n, k \in \mathbb{N}$ we have that

$$\binom{1}{k} + \binom{2}{k} + \dots + \binom{n}{k} = \binom{n+1}{k+1};$$

prove this by fixing an integer $k$ and using induction on $n$.

**Exercise 0.8.** Let $k \in \mathbb{N}$ (i.e., $k$ is a positive integer) and $A_1, \dots, A_k$ be finite sets.

0.8(a) Prove that

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|.$$

[Hint: each element of $x \in A_1 \cup A_2$ is counted once on the LHS (left-hand side). What about the RHS (right-hand side)? You may need to consider a few cases.]

0.8(b) Prove that

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|.$$

0.8(c) Prove that for any $m \in \mathbb{N}$ that

$$\sum_{j=0}^{m} \binom{m}{j} (-1)^j = \binom{m}{0} - \binom{m}{1} + \binom{m}{2} + \dots + (-1)^m \binom{m}{m} = 0$$

[Hint: you may use induction on $m$, or you may use the binomial theorem $(x+y)^m = \sum_{j=0}^{m} x^{m-j} y^j \binom{m}{j}$ and cleverly choose $x, y$.]

0.8(d) Show that

$$|A_1 \cup A_2 \cup \ldots \cup A_k| = \sum_{0\le i\le k} |A_i| - \sum_{0\le i_1<i_2\le k} |A_{i_1} \cap A_{i_2}|$$

$$+ \sum_{0\le i_1<i_2<i_3\le k} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| \cdots + (-1)^k |A_1 \cap A_2 \cap \ldots \cap A_k|.$$

[Hint: you can use part (3) of this exercise; or you can ignore part (3) and use induction.]

**Exercise 0.9.** Let $a_n$ be the number of strings of length $n$ in $\{0,1\}$ such that each 0 must immediately follow and immediately precede a 1; examples of such strings would be 11010111 and 1111, but not 110 or 1001.

0.9(a) Write out the values of $a_n$ for $n = 1, \ldots, 6$.
0.9(b) Prove that $a_n = a_{n-1} + a_{n-2}$ for all $n \ge 3$.
0.9(c) What is the capacity of $\{a_n\}$.
0.9(d) Show that $\{a_n\}$ is a walk count: draw or describe the graph and the set of beginning and of ending vertices.

**Exercise 0.10.** Let $a_n$ be the number of strings of length $n$ in $\{0,1\}$ such that each 0 must immediately follow and immediately precede a 1; examples of such strings would be 11010111 and 1111, but not 110 or 1001.

0.10(a) Write out the values of $a_n$ for $n = 1, \ldots, 6$.
0.10(b) Prove that $a_n = a_{n-1} + a_{n-2}$ for all $n \ge 3$.
0.10(c) What is the capacity of $\{a_n\}$.
0.10(d) Show that $\{a_n\}$ is a walk count: draw or describe the graph and the set of beginning and of ending vertices.

## 1. Least Squares Curve Fitting

Least squares curve fitting is used to fit a model with some parameters to a set of data points. A special case of this is often called *linear regression*, which we now describe. This section also follows the style of Prof. Klaus Hoechsmann, in that it mostly discusses only $2 \times 2$ linear systems.

1.1. **Linear Regression.** Say that we are given "data" points $(x_1, y_1), \ldots, (x_n, y_n)$ in the plane that do not lie on any line $y = a + bx$; we want to find an $a, b$ that is "closest" to exactly fitting these data points: more precisely, we define the *squared error* of a line $y = a + bx$ to these data points to be

$$E(a,b) \stackrel{\text{def}}{=} \sum_{i=1}^{n} (y_i - a - bx_i)^2,$$

and the *least squares fit* to be a pair $a, b \in \mathbb{R}$ that minimize $E(a,b)$. We emphasize that here the $x_i$ and $y_i$ are fixed, and $a, b$ are variables. Note that $E(a,b) \ge 0$ for all $a, b$, and $E(a,b) = 0$ implies that $y_i = a + bx_i$ for all $i$ (in which case $y = a + bx$ is an "exact fit" to the data points). It is well known that $E(a,b)$ is minimized at

the point $(a, b) = (a^*, b^*)$ iff

$$na^* + \left(\sum_{i=1}^{n} x_i\right) b^* = \left(\sum_{i=1}^{n} y_i\right) \tag{7}$$

$$\left(\sum_{i=1}^{n} x_i\right) a^* + \left(\sum_{i=1}^{n} x_i^2\right) b^* = \left(\sum_{i=1}^{n} x_i y_i\right) \tag{8}$$

1.2. **Remarks on the Formulae for Linear Regression.** I discovered (7) and (8) in a list of formulas at the back of a scientific calculator manual while in high school (before scientific calculators were common), at a time when science classes would ask us to fit straight lines to data points by "drawing a line that looked reasonable." Instead I used the above formulas; I also wondered why they held (since they were in the back of a manual with no explanation, before the days of the internet) and whether they held in greater generality. For example, you can probably guess how to generalize these formulas from fitting with lines $y = a + bx$ to fitting with parabolas $y = a + bx + cx^2$ (useful to estimate the acceleration due to the Earth's gravity). In Math 223 we will learn that almost any reasonable guess for a generalization of the above linear regression equations turn outs out to work (i.e., to be true).

1.3. **Linear Regression and $2 \times 2$ Systems.** It is not hard to check by substitution that any $2 \times 2$ system, i.e., system of two equations with two unknowns $a, b$,

$$\alpha a + \beta b = \gamma$$
$$\delta a + \epsilon b = \zeta$$

(where $\alpha, \ldots, \zeta$ are given) has a unique solution iff $\alpha\epsilon - \beta\delta = 0$ (regardless of the values of $\gamma, \zeta$). Hence (7) and (8) have a unique solution iff

$$\left(\sum_{i=1}^{n} x_i\right)^2 \neq n \left(\sum_{i=1}^{n} x_i^2\right) \ ;$$

we later see that, by the *Cauchy-Schwarz inequality*, this condition fails to hold when and only when $x_1 = x_2 = \ldots = x_n$; this fact should make sense intuitively (exercise).

We will later see that a system of $n \times n$ equations has a unique solution iff the *determinant* of the *matrix of coefficients* is nonzero, and that

$$\det \begin{bmatrix} \alpha & \beta \\ \delta & \epsilon \end{bmatrix} = \alpha\epsilon - \beta\delta$$

is the special case of the $2 \times 2$ system above.

1.4. **Manipulating the Data.** The remarkable fact about the above least squares fit is that one you have computed the coefficients of (7) and (8), i.e.,

$$n, \ \sum_{i=1}^{n} x_i, \ \sum_{i=1}^{n} x_i^2, \ \sum_{i=1}^{n} y_i, \ \sum_{i=1}^{n} x_i y_i,$$

it is easy to remove data points (e.g., points that are "outliers," perhaps from erroneous measurements) and add new ones; hence if you want to manipulate the

data a bit, one can easily compute the new coefficients above and solve the new $2 \times 2$ system.

1.5. **Related Measurements of Fit.** By contrast, it is sometimes more desirable to find the best fit where $E(a,b)$ is replaced with

$$E_1(a,b) \stackrel{\text{def}}{=} \sum_{i=1}^{n} |y_i - a - bx_i|$$

or

$$E_\infty(a,b) = E_{\max}(a,b) \stackrel{\text{def}}{=} \max_{i=1,\dots,n} |y_i - a - bx_i| \ ;$$

these can be solved with linear programming, but (1) this is a more difficult computation, and (2) one cannot add or delete data points as easily as for $E(a,b)$. The least squares fit has a number of other desirable properties, such as having (1) an interpretation in terms of a simple of projection (see below) with nice properties, and (2) a simple way to compute *intervals of confidence* for the values of $a, b$ when the data points are viewed as measurements of points that are truly on a line but subject to errors in measurement. For this reason, the least squares is more popular that the best $E_\infty$ fit (a.k.a. Chebyshev fit) or best $E_1$ fit.

1.6. **Completing the Square Versus Differentiation.** In the next subsection we will use the fact that if $\alpha, \beta, \gamma \in \mathbb{R}$ with $\alpha > 0$, then the function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(a) = \alpha a^2 + \beta a + \gamma$ attains its minimum value when and only when $a = \beta/(2\alpha)$; this can be seen by writing

$$f(a) = \alpha\big(a - \beta/(2\alpha)\big)^2 + C, \quad C = \gamma - \beta^2/(4\alpha)$$

where $C$ is a constant depending on $\alpha, \beta, \gamma$, or by differentiating $f$, i.e., computing $f'(a) = 2\alpha a + \beta$ and setting $f'(a) = 0$ (and realizing that $f$ is a parabola "facing upwards").

We will also allude to the related fact that

$$g(a,b) \stackrel{\text{def}}{=} \alpha a^2 + \beta ab + \gamma b^2$$

is non-negative for all $a, b \in \mathbb{R}$ if $\alpha > 0$ and $C$ above is non-negative (we call such a purely quadratic function *positive semidefinite*, and later in Math 223 we will study such quadratic functions systematically).

1.7. **Derivation of Best Fit.** We now derive (7) and (8) based on some assumptions. First notice that

$$E(a,b) \stackrel{\text{def}}{=} \sum_{i=1}^{n} (y_i - a - bx_i)^2 = a^2 n - ab2\sum x_i + b^2 \sum x_i^2 - a2\sum y_i - b2\sum y_i x_i + C$$

where $C = \sum y_i^2$ is a constant independent of $a, b$ and hence unimportant when minimizing $E$. Now assume that $(a^*, b^*)$ is a local minimum of $E(a,b)$. Then the function $f(a) = E(a, b^*)$ has a local minimum at $a^*$, and hence

$$f'(a) = 2na - 2\sum x_i b^* - 2\sum y_i$$

satisfies $f'(a^*) = 0$ (alternatively one can avoid differential calculus by completing the square, as done in the previous subsection). This implies (7). One similarly considers the function $g(b) = E(a^*, b)$ and derives (8).

The assumption that $E(a,b)$ has a minimum value is not justified above; we will understand this when we discuss *quadratic forms*; however, in the above case, where we have two variables $a, b$, the quadratic part of $E(a,b)$ is

$$a^2 n - ab2\sum x_i + b^2 \sum x_i^2 \ ;$$

and it is not hard to see, by scratch, that this quadratic form is necessarily *positive semidefinite*; the general theory of quadratic forms will tell us that $E(a,b)$ has a minimum value that is attained either at a single point or a line in $\mathbb{R}^2$.

1.8. **Least Squares as a Projection.** When we discuss *projections*, we will interpret the linear regression above as projecting the vector

$$\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

(in $\mathbb{R}^n$) onto the line or plane in $\mathbb{R}^n$ spanned by the vectors

$$\vec{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \ \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

In this case (7) and (8) are viewed as the *normal equations*

$$(\vec{1}\cdot\vec{1})\, a + (\vec{1}\cdot\vec{x})\, b = \vec{1}\cdot\vec{y}$$
$$(\vec{x}\cdot\vec{1})\, a + (\vec{x}\cdot\vec{x})\, b = \vec{x}\cdot\vec{y}$$

We will understand that these equations have a unique solution iff $\vec{1}$ and $\vec{x}$ are *linearly independent*, in which case $\vec{1}$ and $\vec{x}$ span a plane; otherwise $\vec{1}$ and $\vec{x}$ span a line, and the above equations in $a, b$ have infinitely many solutions.

1.9. **More General Least Squares.** The great news about linear regression is that all this generalizes to any way of fitting data points $(x_1, y_1), \ldots, (x_n, y_n)$ to a general linear model

$$y = a_1 f_1(x_1) + \cdots a_m f_m(x_m)$$

where $f_1 = f_1(x), \ldots, f_m = f_m(x)$ are arbitrary (!) functions of $x$. This includes models such as

$$y = a + bx + cx^2, \ y = a + be^x, \ y = a + b\sin(x) + c\cos(x), \ \text{etc.}$$

Similarly, if we have data points $(x_1, y_1, z_1), \ldots, (z_n, x_n, y_n)$ and we model $z$ as a function of $x, y$,

$$z = a_1 g_1(x,y) + \cdots + a_m g_m(x,y)$$

the least squares theory goes through to find $a_1, \ldots, a_m$ that minimize

$$E(a_1, \ldots, a_m) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \big(z_i - a_1 g_1(x_i, y_i) - \cdots - a_m g_m(x_i, y_i)\big)^2.$$

## 2. Linear Algebra Without Linear Algebra

In Math 223 we will occasionally apply "linear algebra without linear algebra." A good example is the following:

(1) it is easy to show that any polynomial of degree $n$ that has at least $n+1$ distinct real roots must be the zero polynomial (by factoring the polynomial or using Rolle's Theorem; exercise);
(2) in view of (1), the principles of linear algebra imply that any $n+1$ data points $(x_0, y_0), \dots, (x_n, y_n)$ with distinct $x_i$ can be fitted with a unique polynomial of degree $n$.

We call this "linear algebra without linear algebra" because we are reaching a conclusion (based on theorems in linear algebra) without seeming to make an actual computation (with the numbers or matrices involved in the linear equations involved).

The above theorem about fitting with polynomials is called Lagrange interpolation and there is a lot to say about this; in particular, Lagrange interpolation is a good example of models that behave poorly due to having too many parameters).

Let us describe this principle in a little more detail.

2.1. $n \times n$ **Systems.** In Math 223 we will learn that the following generalization of Subsection 1.3 holds: for any system of $n$ linear equations in $n$ unknowns,

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \ddots & & \vdots & = & \vdots \\ a_{n1}x_1 & + & a_{n2}x_2 & + & \cdots & + & a_{nn}x_n & = & b_n \end{array}$$

the question of whether or not the above system has a unique solution depends only on the $a_{ij}$ and not on the $b_i$.

[The specific condition on the $a_{ij}$ is not important to us in this article, but is of fundamental importance in Math 223 and can be stated in many equivalent ways, involving a *determinant* or *invertibility* or *rank* or *kernel* (etc.) of an associated *matrix*.]

In class we will give some $2 \times 2$ systems as examples, such as

$$\begin{aligned} x + 2y &= b_1 \\ x + 3y &= b_2 \end{aligned}$$

which has a unique solution for any $b_1, b_2 \in \mathbb{R}$, and

$$\begin{aligned} x + 2y &= b_1 \\ 2x + 4y &= b_2 \end{aligned}, \qquad \text{e.g.} \qquad \begin{aligned} x + 2y &= 5 \\ 2x + 4y &= 10 \end{aligned}, \qquad \begin{aligned} x + 2y &= 5 \\ 2x + 4y &= 11 \end{aligned}$$

which, depending on the constants, either have no solutions or infinitely many.

2.2. **Exact Polynomial Fitting.** If we seek to fit $n+1$ data points $(x_0, y_0), \dots, (x_n, y_n)$ with a polynomial of degree $n$, i.e., with $y = a_0 + a_1 x + \cdots + a_n x^n$, then this amounts to $n+1$ equations in the $n+1$ variables $a_0, \dots, a_n$, specifically

$$a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_n x_i^n = y_i$$

for $i = 0, \dots, n$. Notice that here the $a_i$ are variables and the $x_i, y_i$ are given, which makes things look a bit strange in terms of our usual notation for linear equations.

A similar remark holds for any set of $n+1$ data points to be fitted with a function $y = a_0 f_0(x) + \cdots + a_n f_n(x)$ for arbitrary functions $f_0, \ldots, f_n$, but the constraints on the $x_i$ may be more complicated than that they merely be distinct (depending on $f_0, \ldots, f_n$): indeeed, consider $f_j(x) = \sin(2\pi j x)$ and $x_0, \ldots, x_n$ being any collection of distinct integers; then $f_j(x_i) = 0$ for all $i, j$.

2.3. **Parabola Fitting Example.** Say we are trying to find a parabola through the points $(2,3), (4,5), (6,9)$; i.e., we are looking for a polynomial of degree at most two, $p(x) = a_0 + a_1 x + a_2 x^2$ with $a_0, a_1, a_2 \in \mathbb{R}$ and

$$p(2) = 3,\ p(4) = 5,\ p(6) = 9.$$

This amounts to the linear equations:

$$\begin{array}{ccccccc} a_0 & + & 2a_1 & + & 2^2 a_2 & = & 3 \\ a_0 & + & 4a_1 & + & 4^2 a_2 & = & 5 \\ a_0 & + & 6a_1 & + & 6^2 a_2 & = & 9 \end{array}$$

The theorem on linear systems (that we will prove later in this course) tells us that there *exists* a *unique* solution iff (if and only if) the *corresponding homogeneous system*

$$\begin{array}{ccccccc} a_0 & + & 2a_1 & + & 2^2 a_2 & = & 0 \\ a_0 & + & 4a_1 & + & 4^2 a_2 & = & 0 \\ a_0 & + & 6a_1 & + & 6^2 a_2 & = & 0 \end{array}$$

has a unique solution (which would have to be $a_0 = a_1 = a_2 = 0$). But any solution to the homogeneous system means that $q(x) = a_0 + a_1 x + a_2 x^2$ would have three distinct roots, namely $x = 2, 4, 6$; the only polynomial with three distict roots is the zero polynomial.

2.4. **Calculus Example.** Formulas for indefinite integrals look like

$$\int x^2\, dx = \frac{x^3}{3} + C, \quad \int \cos(x)\, dx = \sin(x) + C$$

where $C$ is a constant. The reason that we need to add a $C$ can be viewed with an analogue to *homogeneous systems*. Let us give an example.

The function $f(x) = (x^3/3) + 100$ is a solution of the equation

$$\frac{d}{dx}\, f = x^2.$$

Given one solution $f(x) = (x^3/3) + 100$ to the above equation, we can find all other solutions, $g = g(x)$, as follows: if

$$\frac{d}{dx}\, g = x^2 = \frac{d}{dx}\, f$$

then

$$\frac{d}{dx}(g - f) = 0.$$

We easily see that if $h = h(x)$ satisfies

$$\frac{d}{dx} h = 0$$

then $h = C$ is a constant. It then follows that $g - f$ above equals a constant. Hence given a solution $f = f(x)$ to any "linear equation"

$$\frac{d}{dx}\, f = x^2,$$

all other solutions are given as $f(x) + C$.

Note that it may not be so easy *to solve* a differential equation, such as

$$\frac{d}{dx}\, f = \sin(4x+3)\, \sqrt{x^3 - 5}\, e^{e^{1/x}} + \cos\bigl(1/(x-4)^2\bigr);$$

so while the *existence* of a solution to a differential equation may not be easy to find, the question of *uniqueness* or, more generally, finding all solutions from a single solution tends to be a much easier question, at least if the equation is *linear*.

2.5. **The Uniqueness-Homogeneous Principle.** Rather than begin with the definition of a *real vector space* (e.g., page 15 of the textbook by Jänich), we want to see why we need them. To present an abstraction of the principle of homogenous equations above, we consider a map of sets

$$\mathcal{L}\colon S \to T\ .$$

For example

(1) $S = T = \mathbb{R}^2$ (i.e. the set of pairs of real numbers), and $\mathcal{L}$ takes $(x, y)$ to $(x + 2y, 2x + 4y)$;
(2) $\mathcal{D}$ is the *difference operator*, taking a function $f\colon \mathbb{N} \to \mathbb{Z}$ to the function $\mathcal{D}f$ given by

$$(\mathcal{D}f)(n) = f(n+1) - f(n); \tag{9}$$

here $S = T$ is the set of functions $\mathbb{N} \to \mathbb{Z}$;
(3) the operator $\mathcal{D}$ defined by (9), but acting on $S = T$ on the functions $\mathbb{Z} \to \mathbb{Z}$, $\mathbb{Z} \to \mathbb{R}$, etc.; $\mathcal{D}$ as it acts on $S = T = \mathrm{Poly}_{\le 3}(\mathbb{R})$ (denoted $\mathcal{P}_3$ in the textbook by Jänich) or from $S = \mathrm{Poly}_{\le 3}(\mathbb{R})$ to $T = \mathrm{Poly}_{\le 7}(\mathbb{R})$ or $T = \mathrm{Poly}_{\le 2}(\mathbb{R})$ (pedantically, these various instances of $\mathcal{D}$ are different because their domain and/or codomains are different);
(4) the operator $d/dx$ (differentiation) from $S$ being the differentiable functions $\mathbb{R} \to \mathbb{R}$ to the set $T$ of functions $\mathbb{R} \to \mathbb{R}$
(5) the examples in the exercises below;
(6) we will not allow *non-linear* functions $\mathcal{L}$ between *linear (or vector) spaces* such as $S, T$ in the above examples.

When we solve an equation

$$\mathcal{L}(s) = \text{some given element of } T,$$

we want to determine if such a solution $s$—if it exists—is unique; more generally, given a solution $s$, we want to know all solutions to the above equation. Let us proceed carefully.

We want to know when $s_1, s_2 \in S$ satisfy

$$\mathcal{L}(s_1) = \mathcal{L}(s_2),$$

which is an equation in $T$ (perhaps we know $s_1$ and we want to find all $s_2$ that satisfy this equation, but this is not important here). We wish to have a "subtraction in $T$" that allows us to write

$$\mathcal{L}(s_1) - \mathcal{L}(s_2) = \mathcal{L}(s_2) - \mathcal{L}(s_2),$$

and we want to be able to rewrite this as

$$\mathcal{L}(s_1 - s_2) = \mathcal{L}(s_1) - \mathcal{L}(s_2) = \mathcal{L}(s_2) - \mathcal{L}(s_2) = 0_T,$$

which requires having (1) a zero element of $T$, $0_T$, with respect subtraction, (2) a "subtraction in $S$," and (3) $\mathcal{L}$ has to respect subtraction in $S$ and $T$ in the sense that

$$\mathcal{L}(s_1 - s_2) = \mathcal{L}(s_1) - \mathcal{L}(s_2).$$

Given all this we define the *kernel* or *nullspace* of $\mathcal{L}$ to be

$$\ker(\mathcal{L}) = \{s \in S \mid \mathcal{L}(s) = 0_T\},$$

which is the analogue of the "homogeneous form of system of equations," and we have

$$\mathcal{L}(s_1) = \mathcal{L}(s_2) \iff s_1 - s_2 \in \ker(\mathcal{L}).$$

To summarize, the above relies on subtraction operations in $S, T$ that work in the usual way, and the assumption that $\mathcal{L}$ *respects* this operation.

In *linear algebra over* $\mathbb{R}$, we make the additional assumption that the elements of $S, T$ can be *multiplied (or scaled)* by real numbers, and that $\mathcal{L}$ respects also scaling, i.e.,

$$\mathcal{L}(\alpha s) = \alpha \mathcal{L}(s)$$

for $\alpha \in \mathbb{R}$ and $s \in S$; again the above "respecting" condition involves scaling by $\alpha$ in $S$ (i.e., $\alpha s$), and scaling by $\alpha$ in $T$ (i.e., $\alpha \mathcal{L}(s)$). For example, we have

$$\frac{d}{dx}\big(201.9 f(x)\big) = 201.9 \frac{d}{dx}\big(f(x)\big),$$

and, more generally, for any $\alpha \in \mathbb{R}$ we have

$$\frac{d}{dx}\big(\alpha f(x)\big) = \alpha \frac{d}{dx}\big(f(x)\big).$$

Introducing scaling (or *scalar multiplication*) by real numbers has many advantages. This also gives us an "addition operation"

$$s_1 + s_2 \stackrel{\text{def}}{=} s_1 - \big((-1)s_2\big) \tag{10}$$

which is often convenient. Furthermore, if $s_1, s_2 \in \ker(\mathcal{L})$ as above, then also $\alpha s_1 + \beta s_2 \in \ker(\mathcal{L})$ for any $\alpha, \beta \in \mathbb{R}$.

More generally, one can work with linear algebra over $\mathbb{F}$, where $\mathbb{F}$ is something other than $\mathbb{R}$, such as $\mathbb{N}, \mathbb{Z}, \mathbb{C}, \mathbb{Q}, \ldots$. Here are some considerations for what we need from $\mathbb{F}$:

(1) generally we want a multiplication and addition, including a $-1$ element that functions as in (10);
(2) to speak of *bases*, *basis exchange*, and (*Gaussian elimination* to solve) *equations*, we need to be able to divide by any non-zero element;
(3) when we work with *projections*, we need to be able to take square roots;
(4) when we work with *eigenvalues*, we sometimes need to have any polynomial with coefficients in $\mathbb{F}$ to have all its roots "in $\mathbb{F}$."

In practice we will take $\mathbb{F} = \mathbb{R}$, which has all the above properties except the last one; when we need the last one, we will pass from $\mathbb{R}$ to $\mathbb{C}$; $\mathbb{R}$ also has the advantage that it is a natural setting for many applications.

**Exercise 2.1.** Let $\mathcal{L}$ be the map taking a differentiable function, $f$, to the function $\mathcal{L}f$ defined by

$$\mathcal{L}f = \frac{d}{dx}f - 3f.$$

2.1(a) Show that for any $C \in \mathbb{R}$, $f(x) = Ce^{3x}$ lies in $\ker(\mathcal{L})$.
2.1(b) Show that if $f \in \ker(\mathcal{L})$, then $g(x) \stackrel{\text{def}}{=} f(x)e^{-3x}$ satisfies $g'(x) = 0$ for all $x$.
2.1(c) Show that if $f \in \ker(\mathcal{L})$, then $f(x)$ must be of the form $Ce^{3x}$ for some $C \in \mathbb{R}$.
2.1(d) Find a polynomial of degree one, $p(x) = a_0 + a_1 x$, such that $\mathcal{L}p = x$.
2.1(e) Find all solutions to the equation $\mathcal{L}f = x$.

**Exercise 2.2.** Let $\mathcal{L}$ be the map taking a function $f\colon \mathbb{Z} \to \mathbb{R}$ to the function $\mathcal{L}$ defined by

$$(\mathcal{L}f)(n) = f(n+1) - 2f(n).$$

Show that $f \in \ker(\mathcal{L})$ iff $f$ is given as

$$f(n) = C2^n$$

for some $C \in \mathbb{R}$.

**Exercise 2.3.** Let $\mathcal{L}_{\text{Fib}}$ be the map taking a function $f\colon \mathbb{Z} \to \mathbb{R}$ to the function $\mathcal{L}_{\text{Fib}}f$ defined by

$$(\mathcal{L}_{\text{Fib}}f)(n) = f(n+2) - f(n+1) - f(n).$$

2.3(a) Let $F\colon \mathbb{Z} \to \mathbb{R}$ be the *Fibonacci numbers*, given by
(a) $F(1) = F(2) = 1$,
(b) $F(n) = F(n-1) + F(n-2)$ for $n \geq 3$,
(c) $F(n-2) = F(n) - F(n-1)$ for $n \leq 0$,
which yields the familiar sequence

$$\ldots 13, -8, 5, -3, 2, -1, 1, 0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots$$

Show that $F \in \ker(\mathcal{L}_{\text{Fib}})$.
2.3(b) Show that for any $r \in \mathbb{R}$, the function $g\colon \mathbb{Z} \to \mathbb{R}$ given by $g(n) = r^n$ lies in $\ker(\mathcal{L}_{\text{Fib}})$ iff $r$ satifies

$$r^2 - r - 1 = 0.$$

2.3(c) Let $\xi_+ = (1 + \sqrt{5})/2$ and $\xi_- = (1 - \sqrt{5}/2$. Show that for any $b_0, b_1 \in \mathbb{R}$ there are unique $x, y$ with

$$\begin{aligned} x \;+\; y \;&= b_0 \\ \xi_+ x + \xi_- y \;&= b_1 \end{aligned}$$

2.3(d) Explain why every element, $f$, of $\ker(\mathcal{L}_{\text{Fib}})$ is uniquely determined by its values $f(0)$ and $f(1)$.
2.3(e) Explain why every element, $f$, of $\ker(\mathcal{L}_{\text{Fib}})$ is uniquely expressible as

$$f(n) = x\,\xi_+^n + y\,\xi_-^n$$

for some $x, y \in \mathbb{R}$ [Hint: show that there is a unique $x, y$ satisfying this formula for $n = 0$ and $n = 1$.]
2.3(f) Find a formula for the Fibonacci numbers, $F(n)$, with $n$ above.

**Exercise 2.4.** Let $\mathcal{L}$ be the operator taking a function $f\colon \mathbb{Z} \to \mathbb{R}$ to the function $\mathcal{L}f$ defined by

$$(\mathcal{L}f)(n) = f(n+2) - f(n).$$

2.4(a) Show that $f \in \ker(\mathcal{L})$ iff $f$ is of the form

$$f = \begin{cases} a, & \text{if } n \text{ is even, and} \\ b, & \text{otherwise.} \end{cases}$$

for some $a, b \in \mathbb{R}$.

2.4(b) Show that for any $r \in \mathbb{R}$, the function $g\colon \mathbb{Z} \to \mathbb{R}$ lies in $\ker(\mathcal{L})$ iff $r$ satifies

$$r^2 - 1 = 0.$$

2.4(c) Show that $f \in \ker(\mathcal{L})$ iff $f$ is of the form

$$f(n) = x + (-1)^n y$$

for some $x, y \in \mathbb{R}$.

**Exercise 2.5.** Let $\mathcal{L}$ be the operator taking a function $f\colon \mathbb{Z} \to \mathbb{R}$ to the function $\mathcal{L}f$ defined by

$$(\mathcal{L}f)(n) = f(n+4) - f(n).$$

2.5(a) Say that $f \in \ker(\mathcal{L})$ and $f(0) = 0$, $f(1) = 1$, $f(2) = 2$, and $f(3) = 3$. Describe $f(n)$ for all $n$.

2.5(b) Show that for any $r \in \mathbb{C}$, the function $g\colon \mathbb{Z} \to \mathbb{R}$ lies in $\ker(\mathcal{L})$ iff $r$ satifies

$$r^4 - 1 = 0.$$

2.5(c) Show that the solutions to $r^4 - 1 = 0$ are given by $r = 1, -1, i, -i$ where $i \in \mathbb{C}$ denotes $\sqrt{-1}$.

2.5(d) Show that $f \in \ker(\mathcal{L})$ iff $f$ is of the form

$$f(n) = \alpha + \beta i^n + \gamma(-1)^n + \delta(-i)^n$$

for some $\alpha, \beta, \gamma, \delta \in \mathbb{R}$.

## 3. Sums of Powers

It is well known that

$$1 + 2 + \cdots + n = \binom{n+1}{2}, \tag{11}$$

$$1 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}; \tag{12}$$

similarly one also has

$$1 + 2^k + \cdots + n^k = p_k(n) \tag{13}$$

where $p_k$ is a polynomial in $n$ of degree $k+1$ (there are a number of ways to prove this, see the exercises and the discussion below). It is not hard to see that $p_k(n)$ is divisible by $n(n+1)$ and has leading term $n^{k+1}/(k+1)$ (which is equivalent to the fact that $\int x^k = x^{k+1}/(k+1) + C$). The exact formula is a bit of a mess, but is well studied (type "Bernoulli numbers" into an internet search engine); in this section we give one way to determine $p_k(n)$.

3.1. **Easy Derivation of the Sums of Squares Formula.** In this subsection we wish to outline an easy way to derive the sum of squares formula (12), at least a *conceptually easy* way. We also point out that (12) is connected to the origins of calculus over 2,000 years ago and Archimedes' *quadrature of the parabola* (although Archimedes used a different technique); the point being that from (12) one easily sees that $\int x^2\,dx = x^3/3 + C$.

We claim that if

$$f(n) \stackrel{\text{def}}{=} \sum_{i=1}^{n} k^2, \tag{14}$$

then

(1) $f(n)$ can be written as a polynomial of degree three, i.e., there is a degree three polynomial, $p(x) = a_0 + a_1 x + a_3 x^2 + a_3 x^3$, with $a_0, \ldots, a_3 \in \mathbb{R}$ and $a_3 \neq 0$ such that $f(n) = p(n)$ for all $n \in \mathbb{Z}$; furthermore $a_3 = 1/3$; and
(2) the polynomial $p$ has roots at $x = 0, -1/2, -1$.

Once we have proven these claims, then by high school math,

$$p(x) = (1/3)x(x + 1/2)(x + 1) = x(x + 1)(2x + 1)/6$$

which proves (14). There are a number of subtle aspects of these claims, but once you get used to them (say, by taking Math 223), these ideas will become intuitive.

The idea behind the Claim 2 is to consider to consider values of $f(n)$ for $n$ *negative and zero*, and for $n$ *at the half-integer* $-1/2$. We usually think of $f(n)$ as the sequence

$$f(1) = 1^2 = 1,\ f(2) = 1^2+2^2 = 5,\ f(3) = 1^2+2^2+3^2 = 14,\ f(4) = 30,\ f(5) = 55,\ f(6) = 91,\ \ldots$$

but there is no reason not to extend this sequence *backwards*: since $f(n)$ is given by the recurrence

$$f(n + 1) = f(n) + (n + 1)^2, \quad f(1) = 1,$$

we can use $f(n) = f(n + 1) - (n + 1)^2$ to compute

$$\begin{aligned} f(0) &= f(1) - 1^2 = 1 - 1^2 = 0, \\ f(-1) &= f(0) - 0^2 = 0, \\ f(-2) &= f(-1) - (-1)^2 = 0 - 1 = -1, \\ f(-3) &= f(-2) - (-2)^2 = -1 - 4 = -5, \\ f(-4) &= f(-3) - (-3)^2 = -4 - 9 = -14, \end{aligned}$$

which gives us the two-sided sequence

$$\ldots,\ -14,\ -5,\ -1,\ f(-1) = 0,\ f(0) = 0,\ f(1) = 1,\ 5,\ 14,\ \ldots$$

We remark that this backwards extension of the sum of squares sequence actually follows from (14) and the conventions on summation in Subsection 0.3.

The "backwards extension" of $f$ allows us to view $f$ as a function $\mathbb{Z} \to \mathbb{Z}$, and the calculation of $f(0)$ and $f(-1)$ shows that these are both zero. Furthermore, we easily see that for all $n \in \mathbb{Z}$

$$f(-n) = -f(-1 + n)\ .$$

Assuming that we have proven Claim 1 above,

$$q(x) \stackrel{\text{def}}{=} p(-x) + p(-1 + x)$$

has infinitely many roots/zeros ($q(n) = 0$ for all $n \in \mathbb{Z}$). It is not hard to see that since $p$ is a polynomial, also $q$ is a polynomial (see below); since $q$ has infinitely many roots/zeros $q$ is the zero polynomial. Hence

$$0 = q(-1/2) = p(-1/2) + p(-1/2) = 2p(-1/2)$$

and hence $p(-1/2) = 0$.

It remains to prove Claim 1 above; this can be done in a number of ways, including our discussion in the next subsections.

**Exercise 3.1.** Let $p(x) = a_0 + a_1x + a_3x^2 + a_3x^3$ with $a_3 = 1/3$.

3.1(a) Show that $q(x) = p(-x)$ is also a polynomial. What are its coefficients in terms of $a_0, a_1, a_2$? (Remember that $a_3 = 1/3$).

3.1(b) Show that $q(x) = p(-1 + x)$ is also a polynomial. What are its coefficients in terms of $a_0, a_1, a_2$? (Remember that $a_3 = 1/3$).

3.1(c) Show that $q(x) = p(-x) + p(-1 + x)$ is also a polynomial. What are its coefficients in terms of $a_0, a_1, a_2$? (Remember that $a_3 = 1/3$).

**Exercise 3.2.**

3.2(a) We say that a polynomial $p(x) = a_0 + a_1x + a_3x^2 + a_3x^3$ is *odd* if $p(-x) = -p(x)$. For which $a_0, a_1, a_2, a_3$ is $p$ odd?

3.2(b) Show that if $p(x) = a_0 + a_1x + a_3x^2 + a_3x^3$ is odd, then $p(0) = 0$.

3.2(c) If $p(x) = a_0 + a_1x + a_3x^2 + a_3x^3$, and $q(x) = p(x - 1/2)$ is odd, what can you say about the value of $q(0) = p(-1/2)$? How does this relate to the discussion in this subsection?

3.2(d) We say that a polynomial $p(x) = a_0 + a_1x + a_3x^2 + a_3x^3$ is *even* if $p(-x) = p(x)$. For which $a_0, a_1, a_2, a_3$ is $p$ even?

3.2(e) Show that if $p(x) = a_0 + a_1x + a_3x^2 + a_3x^3$ is even, then $p'(0) = 0$ where $p'$ shorthand for the derivative $dp/dx$.

**Exercise 3.3.** If $f\colon \mathbb{Z} \to \mathbb{R}$ or $f\colon \mathbb{R} \to \mathbb{R}$, we say that

(1) $f$ is *odd* if $f(-x) = -f(x)$ for all $x$ (in the domain of $f$).
(2) $f$ is *even* if $f(-x) = f(x)$ for all $x$ (in the domain of $f$).

3.3(a) Show that if $f\colon \mathbb{Z} \to \mathbb{R}$ or $f\colon \mathbb{R} \to \mathbb{R}$, then

$$f(x) = \frac{f(x) + f(-x)}{2} + \frac{f(x) - f(-x)}{2}$$

expresses $f$ as the sum of an even plus an odd function; in other words, show that the first expression on the RHS (right-hand-side) is an even function, and second expression on the RHS is an odd function, and that the above equation is correct.

3.3(b) If $f\colon \mathbb{Z} \to \mathbb{Z}$, is

$$\frac{f(x) + f(-x)}{2}$$

always a function $\mathbb{Z} \to \mathbb{Z}$? Either (1) show that it is, or (2) give a counterexample or show that it isn't always.

3.3(c) Show that if $f$ is odd, then $f(0) = 0$.

3.3(d) Show that if $f\colon \mathbb{R} \to \mathbb{R}$ is odd and differentiable, then $f' = df/dx$ is even; show the same with "odd" and "even" exchanged.

3.3(e) Show that if $f$ is odd and infinitely differentiable (i.e., has derivatives to all orders), then $f(0), f''(0), f''''(0), \dots$ are zero. Similarly show that if $f\colon \mathbb{R} \to \mathbb{R}$ is even and infinitely differentiable, then $f'(0), f'''(0), \dots$ are zero.

3.3(f) Show any function $\mathbb{Z} \to \mathbb{R}$ or $\mathbb{R} \to \mathbb{R}$ can be expressed *uniquely* as a sum of an even plus an odd function.

**Exercise 3.4.** Say that we know that for any $k \in \mathbb{N}$ there is a polynomial of degree $k+1$, $p_k(n)$, such that

$$\sum_{m=1}^{n} n^k = p_k(n)$$

for all $n \in \mathbb{Z}$. Prove that

3.4(a) for all $k \in \mathbb{N}$, $p_k(x)$ is divisible by $x$ and $x+1$; and

3.4(b) for all even $k \in \mathbb{N}$, $p_k(x)$ is also divisible by $x + 1/2$.

3.2. **Sums of Binomial Coefficients.** Recall that the number of subsets of size $k$ from a fixed set of $n$ elements is

$$\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k(k-1)\dots 1} = \frac{n!}{k!(n-k)!},$$

where ! denotes the "factorial," e.g., $k! = k(k-1)\dots 1$. Equivalently $\binom{n}{k}$ is the number of strings of $x$'s and $y$'s of length $n$ with $k$ $x$'s and $n-k$ $y$'s; this is related to the binomial theorem

$$(x+y)^n = (x+y)(x+y)\dots(x+y) = \sum_{k=0}^{n} x^k y^{n-k} \binom{n}{k}. \tag{15}$$

For this reason the $\binom{n}{k}$ are called *binomial coefficients*. There are many ways to see that

$$\binom{1}{k} + \dots + \binom{n}{k} = \binom{n+1}{k+1}. \tag{16}$$

We remark that you could guess that (16) holds by looking at examples in the first few rows of Pascal's triangle:

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}$$

or, more suggestively, the first few rows of its rotated form:

$$\begin{array}{ccccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \dots \\
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & \dots \\
1 & 3 & 6 & 10 & 15 & 21 & 28 & 36 & \dots \\
1 & 4 & 10 & 20 & 35 & 56 & 84 & 120 & \dots \\
1 & 5 & 15 & 35 & 70 & 126 & 210 & 330 & \dots
\end{array}$$

Let us give one way to see that (16); to simplify the computation, first consider the case $k = 2$: note that

$$\binom{n+1}{3} - \binom{n}{3} = \frac{(n+1)n(n-1)}{2 \cdot 3} - \frac{n(n-1)(n-2)}{2 \cdot 3}$$

$$= n(n-1)\frac{(n+1)-(n-2)}{2\cdot 3} = n(n-1)\frac{3}{2\cdot 3} = \frac{n(n-1)}{2} = \binom{n}{2}.$$

It follows that

$$\binom{1}{2} + \ldots + \binom{n}{2}$$

$$= \left(\binom{2}{3} - \binom{1}{3}\right) + \left(\binom{3}{3} - \binom{2}{3}\right) + \left(\binom{4}{3} - \binom{3}{3}\right) + \cdots + \left(\binom{n+1}{3} - \binom{n}{3}\right)$$

$$= \binom{n+1}{3} - \binom{1}{3} = \binom{n+1}{3}.$$

We can show that (16) holds for any $k \in \mathbb{N}$ similarly, starting by showing that

$$\binom{n+1}{k+1} - \binom{n}{k+1} = \binom{n}{k}$$

because we can write the left-hand-side above as

$$n(n-1)\ldots(n-k+1)\frac{(n+1)-(n-k)}{(k+1)!} = n(n-1)\ldots(n-k+1)\frac{k+1}{(k+1)!} = \frac{n(n-1)\ldots(n-k+1)}{k!}.$$

**Exercise 3.5.** The binomial theorem (15) for $n = 4$ says that

$$(x+y)^4 = (x+y)(x+y)(x+y)(x+y) = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4.$$

Notice that there are four strings with three $x$'s and one $y$:

$$xxxy,\ xxyx,\ xyxx,\ yxxx$$

and six strings with two $x$'s and two $y$'s:

$$xxyy,\ xyxy,\ xyyx,\ yxxy,\ yxyx,\ yyxx.$$

Notice that in both cases above we have listed the strings in *lexicographical order*, meaning the order they would appear in a dictionary (if they were words).

3.5(a) List all strings of one $x$ and three $y$'s in lexicographical order.

3.5(b) List all strings of one $x$ and four $y$'s in lexicographical order.

3.5(c) List all strings of two $x$'s and three $y$'s in lexicographical order.

3.5(d) Using your answer to the last part, **describe—IN 15 WORDS OR FEWER—an algorithm** to list all strings of three $x$'s and two $y$'s in lexicographical order; i.e., **do not produce this list**, but instead describe **how** you would take the list you wrote in the last part **as input** and then **output** a list of all strings of three $x$'s and two $y$'s.

3.5(e) Explain how the number of elements in some of your lists above relate to the binomial theorem

$$(x+y)^5 = x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5y^4 + y^5.$$

**Exercise 3.6.** For any $k \in \mathbb{N}$, prove (16) by induction on $n$ (with $n = 1$ as the base case).

**Exercise 3.7.** Prove (16) directly, by noting that its right-hand-size represents the number of strings of $n-k$ $x$'s and $k+1$ $y$'s, and using the fact that each such string begins with some number of $x$'s before it encounters its first $y$.

3.3. **Sums of Powers.** Let us give a method to find $p_k(n)$ in (13) for all $k \in \mathbb{N}$. Let us illustrate this method to find $p_1(n)$ and $p_2(n)$ (which we already know from (11) and (12)). First note that for all $m \in \mathbb{N}$,

$$\binom{m}{1} = m; \tag{17}$$

for $k = 1$ (16) says that

$$\binom{1}{1} + \binom{2}{1} + \cdots + \binom{n}{1} = \binom{n+1}{2} = \frac{(n+1)n}{2},$$

and hence

$$1 + 2 + \cdots + n = \frac{(n+1)n}{2},$$

which implies (11), i.e., that $p_2(n) = (1/2)(n^2 + n)$. Next note that for any $m \in \mathbb{N}$

$$\binom{m}{2} = \frac{m(m-1)}{2} = (1/2)m^2 + (-1/2)m.$$

Summing over $m \in [n]$ and using (16) with $k = 2$, we have

$$\binom{n+1}{3} = (1/2)p_2(n) + (-1/2)p_1(n)$$

which gives us a formula for $p_2(n)$ (since we know $p_1(n)$). Similarly

$$\binom{m}{3} = \frac{m(m-1)(m-2)}{6} = (1/6)m^3 - (1/2)m^2 + (1/3)m$$

implies that

$$\binom{n+1}{4} = (1/6)p_3(n) - (1/2)p_2(n) + (1/3)p_1(n)$$

which gives us a formula for $p_3(n)$ (since we know $p_1(n), p_2(n)$). We similarly can find $p_k(n)$ for any $k \in \mathbb{N}$.

**Exercise 3.8.** Show that $p_k(n)$ is a polynomial of degree $k+1$ whose leading term is $n^{k+1}/(k+1)$. Use this to show that $\int_0^a x^k \, dx = a^{k+1}/(k+1)$ by evaluating a Riemann sum.

3.4. **LINEARITY AND ABSTRACT VECTOR SPACES.** In the previous subsections, and in the subsection that follows, we are implicitly working with basic concepts regarding *linearity* and *abstract vector spaces*. Let us give some rough ideas.

First, there are a large number of *abstract vector spaces*, or simply *vector spaces*, in the background. For example, we viewed the function

$$f(n) = \sum_{m=1}^{n} m^2 \tag{18}$$

as an element of:

(1) the set of functions $f : \mathbb{N} \to \mathbb{N}$, which we denote Functions$(\mathbb{N} \to \mathbb{N})$ or $\mathbb{N}^{\mathbb{N}}$;
(2) the set of functions $f : \mathbb{N} \to \mathbb{R}$, which we denote Functions$(\mathbb{N} \to \mathbb{R})$ or $\mathbb{R}^{\mathbb{N}}$;
(3) the set of functions $f : \mathbb{Z} \to \mathbb{Z}$ (understand summation with the conventions of Subsection 0.3), which we denote Functions$(\mathbb{Z} \to \mathbb{Z})$ or $\mathbb{Z}^{\mathbb{Z}}$; or
(4) the set of functions $f : \mathbb{N} \to \mathbb{R}$, which we denote Functions$(\mathbb{N} \to \mathbb{R})$ or $\mathbb{R}^{\mathbb{N}}$;

we have also worked with *polynomials*, viewed as:

(1) the set of formal expressions $p(x) = a_0 + a_1x + \cdots + a_3x^3$ where $a_0, \ldots, a_3 \in \mathbb{R}$ (i.e., polynomials of degree at most three over $\mathbb{R}$), which we denote $\mathrm{Poly}_3(\mathbb{R})$ or, in the textbook by Jänich on page 82, $\mathcal{P}_3$;
(2) more generally, the set of formal expressions $p(x) = a_0 + a_1x + \cdots + a_nx^n$ where $a_0, \ldots, a_n \in \mathbb{R}$ (i.e., polynomials over $\mathbb{R}$), which we denote $\mathrm{Poly}(\mathbb{R})$ or $\mathcal{P}$;
(3) the set of functions $\mathbb{R} \to \mathbb{R}$ whose values are given by an element of $\mathrm{Poly}_{\leq 3}$;
(4) the set of functions $f\colon \mathbb{N} \to \mathbb{N}$ for which there is a formula $f(n) = p(n)$ where $p$ is a polynomial over $\mathbb{R}$, which can be viewed as
$$\mathrm{Functions}(\mathbb{N}, \mathbb{N}) \cap \mathrm{Poly}\ ;$$
(5) etc.

If these distictions seem a bit pedantic (which they are ...), just recall that we started with a sequence $1, 5, 14, 30, 55, \ldots$ and spoke about its values at negative indices and at the index $-1/2$. The reason all of this works is that a polynomial of degree 3 is determined by its values on $\mathbb{Z}$ or even just on $\mathbb{N}$, or even just on four distinct real values (exercise from high school algebra, or use Rolle's Theorem in calculus); for similar reasons a polynomial is determined by its values at any infinite set of distict real numbers (exercise).

The key to some of what we do is that for a number of the above sets, if $u, v$ are any two elements and $\alpha, \beta \in \mathbb{R}$, then there is a *sensible* notion of a new element $\alpha u + \beta v$ (e.g., adding of functions or polynomials and multiplying a function or polynomial by a real number); we could also limit ourselves to $\alpha, \beta \in \mathbb{Z}$ or $\alpha, \beta \in \mathbb{N}$, but we will eventually see that working with $\mathbb{R}$ has numerous advantages[1]; sometimes it will be better to work in $\mathbb{C}$. A key idea is that of *linear transformations*; for example, if $p = p(x) \in \mathrm{Poly}_3(\mathbb{R})$, then we used a number of maps $\mathcal{L}\colon \mathrm{Poly}_3(\mathbb{R}) \to \mathrm{Poly}_3(\mathbb{R})$ that were *linear* in that
$$\mathcal{L}(\alpha p_1 + \beta p_2) = \alpha\mathcal{L}(p_1) + \beta\mathcal{L}(p_2),$$
including the maps $\mathcal{L}$ given by

(1) $(\mathcal{L}p)(x) \stackrel{\mathrm{def}}{=} p(-x)$,
(2) $(\mathcal{L}p)(x) \stackrel{\mathrm{def}}{=} -p(-x)$,
(3) $(\mathcal{L}p)(x) \stackrel{\mathrm{def}}{=} p(-1 + x)$,
(4) $(\mathcal{L}p)(x) \stackrel{\mathrm{def}}{=} p(-x) + p(-1 + x)$ (which is a sum of two of the linear transformation above),
(5) etc.

The key fact we need about (18) is that there is a $p \in \mathrm{Poly}_{\leq 3}(\mathbb{R})$ such that $f(n) = p(n)$ for all $n$ or, informally, "$f = p$". In the next subsection we prove this.

### 3.5. The Operators $\mathcal{D}, \mathcal{S}$ and Some Operators from Calculus.

**Definition 3.1.** If $f \in \mathrm{Functions}(\mathbb{N} \to \mathbb{Z})$, i.e., $f\colon \mathbb{N} \to \mathbb{Z}$, we define the *difference of $f$* to be the function
$$(\mathcal{D}f)(n) \stackrel{\mathrm{def}}{=} f(n + 1) - f(n);$$
we refer to $\mathcal{D}$ as the *difference operator* on functions $\mathbb{Z} \to \mathbb{Z}$, $\mathbb{Z} \to \mathbb{R}$, $\mathbb{R} \to \mathbb{R}$, etc.

[1] Sometimes working with the rational numbers is good enough; when we want to talk of projections, which invovle an *inner product* (such as the *dot product*), it is simplest to work with $\mathbb{R}$ or $\mathbb{C}$.

Pedantically, the difference operator $\mathcal{D}$ is a different map or *operator* on each of the above set of functions; generally, it does not cause confusion to use the same letter $\mathcal{D}$ for each of these different situations.

**Theorem 3.2.** *Consider the difference operator*

$$(\mathcal{D}f)(n) \stackrel{def}{=} f(n+1) - f(n)$$

*as it acts on functions $f\colon \mathbb{Z} \to \mathbb{R}$. Then for any $f, g$ functions $\mathbb{Z} \to \mathbb{R}$ we have*

*(1) for any $n \in \mathbb{N}$ we have*

$$(\mathcal{D}f)(1) + (\mathcal{D}f)(2) + \cdots + (\mathcal{D}f)(n) = f(n+1) - f(1); \tag{19}$$

*(2) (19) holds for $n \in \mathbb{Z}$, i.e.,*

$$\sum_{m=1}^{n} (\mathcal{D}f)(m) = f(n+1) - f(1) \tag{20}$$

*where we understand this sum for $n \le 0$ according to our conventions (Subsection 0.3);*

*(3) there is a polynomial, $p$, of degree 3 with leading coefficient $1/3$ such that $(\mathcal{D}p)(n) = n^2$;*

*[the above claims represent our way of finding a formula for $1 + 2^2 + \cdots + n^2$, and the claims below expand on this]*

*(4) if $\mathcal{D}f = 0$ (the zero function), then $f(n) = C$ is constant;*

*(5) if $\mathcal{D}f = \mathcal{D}g$, then $f = g + C$ for some constant $C$, meaning that $f(n) = g(n) + C$;*

*(6) if $f(n) = \binom{n}{k+1}$, then $(\mathcal{D}f)(n) = \binom{n}{k}$ (meaning that $\mathcal{D}f$ is the function taking $n$ to $\binom{n}{k}$);*

*(7) if $f, g\colon \mathbb{Z} \to \mathbb{R}$ and $\alpha, \beta \in \mathbb{R}$, then*

$$\mathcal{D}(\alpha f + \beta g) = \alpha \mathcal{D}(f) + \beta \mathcal{D}(g);$$

*(8) if $(\mathcal{D}f)(n) = n^2$ for all $n$, then $f(n)$ is a polynomial of degree 3;*

*(9) more generally, if there is a polynomial $p$ such that $\mathcal{D}f = p$, i.e., $(\mathcal{D}f)(n) = p(n)$ for all $n \in \mathbb{Z}$, then the values of $f$ are given by a polynomial, and the degree of this polynomial is (exactly) one more than that of $f$ provided that $p \neq 0$.*

For the exercises below, we define an operator $\mathcal{S}$ as taking a function $f$ and returning a function $\mathcal{S}f$ defined by

$$(\mathcal{S}f)(n) \stackrel{\text{def}}{=} f(1) + f(2) + \cdots + f(n) = \sum_{m=1}^{n} f(m)$$

(so either $n \in \mathbb{N}$, or if $n \le 0$ and $n \in \mathbb{Z}$ we use the summation conventions of Subsection 0.3). So if $f(n) = n^2$, we have

$$(\mathcal{S}f)(n) \stackrel{\text{def}}{=} 1 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6},$$

or in shorthand:

$$\mathcal{S}(n^2) = \frac{n(n+1)(2n+1)}{6}.$$

We may write (19) and (20) as

$$\big(\mathcal{S}(\mathcal{D}f)\big) = f(n+1) - f(1).$$

This is a discrete analogue of the calculus theorem:

$$\int_{x=a}^{x=b} f'(x)\,dx = f(b) - f(a).$$

**Exercise 3.9.** Compute the function $(\mathcal{D}f)(n)$ for all $n \in \mathbb{N}$:

3.9(a) $f(n) = (n-1)^2$;
3.9(b) $f(n) = (n-1)n(2n-1)/6$;
3.9(c) $f(n) = \binom{n}{4} \stackrel{\text{def}}{=} n(n-1)(n-2)(n-3)/24$;
3.9(d) $f(n) = -(1/3)^{n-1}/2$ and simplify your answer.
3.9(e) Show how (19) and the above computations yield the following formulas:

$$\begin{aligned}
1 + 3 + 5 + \cdots + (2n-1) &= n^2, \\
1 + 2^2 + 3^2 + \cdots + n^2 &= n(n+1)(2n+1)/6, \\
\binom{1}{3} + \binom{2}{3} + \cdots + \binom{n}{3} &= \binom{n+1}{4} \\
(1/3)^1 + (1/3)^2 + \cdots + (1/3)^n &= \frac{1-(1/3)^n}{2},
\end{aligned}$$

In class we remarked that $\mathcal{D}$, like differentiation, reduces the degree of a polynomial by 1; however,

$$\mathcal{D}\binom{x}{12} \quad \text{and} \quad \frac{d}{dx}x^{12}$$

have "simple formula," whereas

$$\mathcal{D}(x^{12}) \quad \text{and} \quad \frac{d}{dx}\binom{x}{12}$$

do not.

3.6. **Changing From Binomials to Powers and Vice Versa.** Here is another way to look at the above method of finding $p_k(n)$. We have nice formulas to find the sum over all $m \in [n]$ of the functions

$$\binom{m}{1}, \ \binom{m}{2}, \ \binom{m}{3}, \ \binom{m}{4} \ \cdots$$

but we are more interested in such formulas for the functions

$$m, \ m^2, \ m^3, \ m^4, \ \ldots$$

Expanding the $\binom{m}{k}$ gives formulas

$$\begin{aligned}
\tbinom{m}{1} &= m, \\
\tbinom{m}{2} &= (-1/2)\,m + (1/2)\,m^2, \\
\tbinom{m}{3} &= (1/3)\,m + (-1/2)\,m^2 + (1/6)\,m^3, \\
\tbinom{m}{4} &= (-1/4)\,m + (11/24)\,m^2 + (-1/4)\,m^3 + (1/24)m^4.
\end{aligned}$$

In Math 223 we will learn a general method for converting the other way (by *inverting* a *lower triangular matrix*)—although this can be done by hand here—to

obtain

$$\begin{aligned} m &= \tbinom{m}{1} \quad , \\ m^2 &= \tbinom{m}{1} + 2\tbinom{m}{2} \quad , \\ m^3 &= \tbinom{m}{1} + 6\tbinom{m}{2} + 6\tbinom{m}{3} \quad , \\ m^4 &= \tbinom{m}{1} + 14\tbinom{m}{2} + 36\tbinom{m}{3} + 24\tbinom{m}{4}. \end{aligned}$$

So to sum over $m \in [n]$ for the functions $m, m^2, m^3, \ldots$, we use the above conversion. So in view of the above table and (16) we have

$$\begin{aligned} \textstyle\sum_{m=1}^n m &= \tbinom{n+1}{2} \quad , \\ \textstyle\sum_{m=1}^n m^2 &= \tbinom{n+1}{2} + 2\tbinom{n+1}{3} \quad , \\ \textstyle\sum_{m=1}^n m^3 &= \tbinom{n+1}{2} + 6\tbinom{n+1}{3} + 6\tbinom{n+1}{4} \quad , \\ \textstyle\sum_{m=1}^n m^4 &= \tbinom{n+1}{2} + 14\tbinom{n+1}{3} + 36\tbinom{n+1}{4} + 24\tbinom{n+1}{5}. \end{aligned} \tag{21}$$

**Exercise 3.10.** Verify that second formula in (21) gives our usual formula for $\sum_{m=1}^n m^2$.

Next we can use the fact that

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$$

(exercise) to write, to convert the above summation formulas to formulas in terms of $\binom{n}{k}$ instead of $\binom{n+1}{k}$; then we can use the expressions for $\binom{n}{k}$ to convert to powers of $n$.

The general idea of expressing the functions $n^k$ in terms of functions $\binom{n}{k}$—more specifically as *linear combinations* of the $\binom{n}{k}$—and vice versa is known in linear algebra as a *change of basis*.

3.7. **Stirling Numbers.** The particular coefficients in writing the $\binom{n}{k}$ (for $k = 1, 2, \ldots$) in terms of $n^k$ and vice versa are essentially known as *Stirling numbers (of the first and second kind*; more specifically, the Stirling numbers are usually defined as translating between the $n^k$ and the functions

$$n^{(k)} \stackrel{\text{def}}{=} \binom{n}{k} k! = n(n-1)\ldots(n-k+1)$$

(some authors write this as $n_{(k)}$ and write $n^{(k)}$ for $n(n+1)\ldots(n+k-1)$). In this case the $n^{(k)}$ can be expressed in terms of the $n^k$ with *integer coefficients* (since we don't divide by $k!$) in a *matrix* with 1's along the *diagonal*, and hence the $n^k$ can be expressed in terms of the $n^{(k)}$ with *integer coefficients* (exercise, using $(I+L)^{-1} = I - L + L^2 - \ldots$, or by induction on $k$).

3.8. **Integrals of Even Powers of** $\cos(x)$**.** Another idea of *change of basis* is a way to integrate

$$\cos^2(x), \cos^4(x), \cos^6(x), \ldots \tag{22}$$

It is easy to integrate the functions

$$\cos(2x), \cos(4x), \cos(6x), \ldots \tag{23}$$

and one can express the functions in (22) as *linear combinations* of those in (23). (exercise)

## 4. Fibonacci Numbers and Recurrence Equations

The Fibonacci are defined by the "initial condition"

$$f(1) = 1,\ f(2) = 1 \tag{24}$$

and the "recurrence equation"

$$f(k) = f(k-1) + f(k-2) \quad \text{for } k \geq 3\ . \tag{25}$$

Fibonacci occur many applications, e.g., Fibonacci heaps, Fibonacci search for the maximum value of a function on an interval, etc. (this "etc." is a very long list). They are also a canonical example of recurrence equations in algorithms, and have been studied by mathematicians in some form for 2000-3000 years (Fibonacci popularized their study in Europe).

For $\alpha, \beta \in \mathbb{R}$ we can generalize the above sequence to the initial conditions

$$f(1) = \alpha,\ f(2) = \beta$$

(keeping the recurrence equation (25)). If we denote the resulting sequence $f_{\alpha,\beta}(k)$ then we have linear relation

$$f_{\alpha,\beta} = \alpha f_{1,0} + \beta f_{0,1}\ ;$$

hence to find a formula for $f_{\alpha,\beta}(k)$ it suffices a formula for $f_{1,0}$ and $f_{0,1}$.

4.1. **Properties of Fibonacci Numbers.** Likely the reader knows of some of the properties of Fibonacci numbers, such as the fact that there is an exact formula and that $f_n$ is the integer nearest to $\xi_+^n/\sqrt{5}$ where $\xi_+$ is the *golden ratio* (Exercises 0.6(f) and 0.6(g)).

4.2. **Solution to the Fibonacci Recurrence by Guessing and Solving a $2\times 2$ System.** One solves (25) by guessing the existence of a solution $f_n = r^n$ (therefore different initial conditions than (24)), which holds for $r$ that satisfies

$$r^{n+2} = r^{n+1} + r^n$$

an equation that holds independent of $n$ for $r$ satisfying

$$r^2 - r - 1 = 0,$$

i.e.,

$$r_+, r_- \stackrel{\text{def}}{=} \frac{1 \pm \sqrt{5}}{2}.$$

For any $f_1, f_2$ there are unique $c_+, c_- \in \mathbb{R}$ for which

$$f_i = c_+ r_+^i + c_- r_-^i$$

for $i = 1, 2$ (this is a $2 \times 2$ system). Hence, by the above guessing and verification procedure, we produce one solution $f_n = c_+ r_+^n + c_- r_-^n$ to (25) with the correct values of $f_1, f_2$; however clearly $f_3$ is uniquely determined by $f_2, f_1$, similarly $f_4, f_5, \ldots$ are uniquely determined by $f_2, f_1$. Hence we have produced this unique solution.

4.3. **Solving General Recurrence Relations.** A general "$k$-th order" recurrence equation

$$a_0 f_n + a_1 f_{n-1} + \cdots + a_k f_{n-k} = 0 \tag{26}$$

with $a_0, a_k \neq 0$ (otherwise we can write a lower order recurrence equation) can be similarly solved, by solving the equation

$$a_0 r^k + a_1 r^{k-1} + \cdots + a_k = 0$$

If $r = \rho$ is a root of order $m$, then $\rho \neq 0$ and there are solutions

$$f_n = \rho^n n^i$$

for $i = 0, \ldots, m-1$. It is easy to verify that these are solutions by introducing the *shift operator*, $\sigma$, on functions $\mathbb{N} \to \mathbb{C}$ (or $\mathbb{Z} \to \mathbb{C}$) defined by

$$(\sigma f)(n) \stackrel{\text{def}}{=} f(n+1).$$

Then (26) is equivalent to

$$\big(a_0 \sigma^k + a_1 \sigma^{k-1} + \cdots + a_k\big) f = 0$$

where we set $f(n) = f_n$ (i.e., regard $f_n$ as a function $\mathbb{Z} \to \mathbb{C}$); furthermore, if

$$a_0 x^k + \cdots + a_k = a_0 (x - r_1) \ldots (x - r_k)$$

(i.e., $r_1, \ldots, r_k$ are the roots of this polynomial, listed with their multiplicities), then if $\rho$ occurs $m$ times among the $r_1, \ldots, r_k$, then

$$(\sigma - \rho)^m f = 0$$

implies (26). Then we verify that if $f(n) = \rho^n p(n)$ where $p$ is a polynomial, then $((\sigma - \rho) f) = \rho^n q(n)$ where $q$ is a polynomial of degree one less than $p$.

**Exercise 4.1.** It is easy to give a natural definition for $P(\sigma)$ for any polynomial $P = P(x)$ over $\mathbb{R}$ or $\mathbb{C}$ (or any *field*) and to verify that for any polynomials $P, Q$

$$P(\sigma) Q(\sigma) = (PQ)(\sigma)$$

where $PQ$ is multiplication of polynomials, and hence

$$P(\sigma) Q(\sigma) = Q(\sigma) P(\sigma).$$

It follows that if $r_1, \ldots, r_k$ are the $k$ roots of

$$p(x) = a_0 x^k + a_1 x^{k-1} + \cdots + a_k,$$

i.e.,

$$p(x) = (x - r_1) \ldots (x - r_k),$$

then

$$p(\sigma) = (\sigma - r_1) \ldots (\sigma - r_k),$$

and when applying $p(\sigma)$ to a function, one is free to permute the factors $(\sigma - r_1), \ldots, (\sigma - r_k)$ as one likes.

4.4. **Recurrences and Matrix Powers.** When we discuss matrices, we will see that (25) can be written in matrix form as

$$\begin{bmatrix} f_n \\ f_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} f_{n-1} \\ f_{n-2} \end{bmatrix}$$

from which we derive the matrix formula

$$\begin{bmatrix} f_n \\ f_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{n-2} \begin{bmatrix} f_2 \\ f_1 \end{bmatrix} .$$

It will turn out that $(1 \pm \sqrt{5})/2$ are the *eigenvalues* of this matrix, from which we can also derive our exact formula for the Fibonacci numbers and, more generally, any solution to (25).

Such remarks are valid for any recurrence of the form (26). It will turn out that the sitation above where $\rho$ is root of order $m \geq 2$ is a *defect* of the matrix analogous to the above matrix, i.e., to the matrix

$$\begin{bmatrix} -a_1/a_0 & -a_2/a_0 & \cdots & -a_{n-1}/a_0 & -a_n/a_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} ;$$

this gives rise to an $m \times m$ *Jordan block* with *eigenvalue* $\rho$ in the above matrix.

4.5. **(Additional) Exercises.**

## 5. Moving Averages (A Bit of Time Series)

Time series is a large field, useful in financial and related forecasting. A basic example is the study of *moving averages*.

Say that $f\colon \mathbb{N} \to \mathbb{R}$ represents the price of a financial instrument traded on a public exchange, with $f(i)$ denoting the price at day $i$. The *30-day moving average* of $f$ is defined as

$$(\mathcal{M}_{30} f)(n) = \frac{f(n) + f(n+1) + \cdots + f(n+29)}{30},$$

which makes $\mathcal{M}_{30}$ an operator on functions $\mathbb{N} \to \mathbb{R}$. For any other $g\colon \mathbb{N} \to \mathbb{R}$ and $\alpha, \beta \in \mathbb{R}$ we have

$$\mathcal{M}_{30}(\alpha f + \beta g) = \alpha \mathcal{M}_{30} f + \beta \mathcal{M}_{30} g. \tag{27}$$

If $f(n) = c$ is a constant function, then $\mathcal{M}_{30} f$ is again the constant function $c$. If $f(n) = (-1)^n$, then $\mathcal{M}_{30} f = 0$. More generally, if $f(n) = \cos(2\pi(n+\phi)/M)$ for any $\psi \in \mathbb{R}$ and integer $M$ greater than 2 and dividing 30 (the special case $M = 2$ and $\phi = 0$ is the function $f(n) = (-1)^n$). Let us try to organize the above ad hoc observations more systematically.

The first statement says that the constant function is *invariant under* $\mathcal{M}_{30}$). The other statements attempt to describe the *kernel of* $\mathcal{M}_{30}$, i.e., the set of functions

$$\ker(\mathcal{M}_{30}) = \{f \mid \mathcal{M}_{30} f = 0\}$$

The linearity of $\mathcal{M}_{30}$, i.e., (27), implies that if $f, g \in \ker(\mathcal{M}_{30})$, then also

$$\alpha f + \beta g \in \ker(\mathcal{M}_{30})$$

for any $\alpha, \beta \in \mathbb{R}$.

More generally, if $\mathcal{M}$ is a *linear map* between *vector spaces*, e.g., $\mathcal{M}\colon \mathbb{R}^n \to \mathbb{R}^m$, then the set

$$\ker(\mathcal{M}) = \{f \mid \mathcal{M}f = 0\}$$

is called the *kernel of* $\mathcal{M}$. If $f, g \in \ker(\mathcal{M})$, then any *linear combination of* $f, g$, i.e., function/vector of the form $\alpha f + \beta g$, is also in the kernel of $\mathcal{M}$.

The kernel $\mathcal{M}_{30}$ are those functions that $\mathcal{M}_{30}$, roughly speaking, "smooths away" (in the moving average). More generally, a function $f$ is in the kernel of $\mathcal{M}_{30}$ iff $f$ satisfies the recurrence equation

$$f(n) + f(n+1) + \cdots + f(n+29) = 0, \tag{28}$$

in which case $f$ satisfies the recurrence

$$1 + \sigma + \cdots + \sigma^{29} = 0.$$

According to Subsection 4.3, we can find the general solution to the above recurrence by finding the solutions

$$1 + r + \cdots + r^{29} = 0; \tag{29}$$

multiplying by $1 - r$ we see that the solutions are precisely those $r$ with $r \neq 1$ but $r^{30} = 1$. Taking $\zeta = e^{2\pi i/30}$, we see that the solutions to (29) are given by

$$\zeta, \zeta^2, \ldots, \zeta^{29},$$

and since these are all distinct we see that the general solution to (28) is given as

$$f(n) = c_1\zeta^n + c_2\zeta^{2n} + \cdots + c_{29}\zeta^{29}.$$

Hence this is the general form of an element in the kernel of $\mathcal{M}_{30}$.

Notice that even if we are interested only in real-valued $f$, it is can be more convenient and intuitive to use complex *30-th roots of unity*, i.e., $\zeta, \ldots, \zeta^{29}$, rather than the equivalent expressions in terms of sines and/or cosines.

Functions $\mathbb{N} \to \mathbb{R}$ turn out to be an *infinite dimensional vector space*. Alternatively, one could view $\mathcal{M}_{30}$, based on 200 days of data, as a map from $\mathbb{R}^{200} \to \mathbb{R}^{171}$ (since the moving average at day 172 requires knowledge of the price on day $172 + 29 = 201$).

## 6. Linearity in Power Series

Although spaces of functions, of power series, etc., tend to be infinite dimensional, they give a very natural motivation for the notion of linearity.

### 6.1. **Trigonometric Functions.** The power series

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$$
$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots$$

converge for any $x \in \mathbb{R}$ and are useful for approximating $\sin(x), \cos(x)$. From these series we get other series for

$$2\sin(x),\ 10\sin(x) + 30\cos(x)$$

in the evident "linear fashion," and more generally, for any $\alpha, \beta \in \mathbb{R}$ we have

$$\alpha \sin(x) + \beta \cos(x) = \beta + \alpha x - \frac{\beta}{2!}x^2 - \frac{\alpha}{3!}x^3 + \cdots$$

There is another, more sutble, form of linearity: the above series assume that $x$ is measured in radians. If you prefer degrees, $y$, then (keeping track of units) $y^\circ = (\pi y/180)^{\text{rad}}$, and so

$$\sin(y^\circ) = \sin\big((\pi y/180)^{\text{rad}}\big) = \frac{\pi}{180}y - \frac{\pi^3}{180^3 \cdot 3!}y^3 + \cdots$$

and, more generally, for any $\gamma \in \mathbb{R}$ we have

$$\sin\big((\gamma x)^{\text{rad}}\big) = \gamma x - \frac{\gamma^3}{3!}x^3 + \cdots \tag{30}$$

6.2. **Taylor Series.** The origin of the above formulas is Taylor's Theorem

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots$$

for sufficiently "nice" functions; alternatively one can truncate the series above to approximate $f(x)$, and Taylor's Theorem allows you to bound the error (if $f$ is sufficiently differentiable).

If $g(x)$ is another sufficiently nice function, we have for any $\alpha, \beta \in \mathbb{R}$

$$\alpha f(x) + \beta g(x) = \big(\alpha f(0) + \beta g(0)\big) + \big(\alpha f'(0) + \beta g'(0)\big)x + \frac{\alpha f''(0) + \beta g''(0)}{2}x^2 + \cdots$$

So there is a "linearity" in combining $f, g$ in this way, which is reflected in the fact that

$$(\alpha f + \beta g)'(0) = \alpha f'(0) + \beta g'(0)$$

and similarly for higer order derivatives.

If rather than scaling $f, g$ one scales $x$, one has that if $h(x) = f(\gamma x)$ for a $\gamma \in \mathbb{R}$, then (by the chain rule)

$$h(0) = f(0),\ h'(0) = \gamma f'(0),\ h''(0) = \gamma^2 f''(0),\ \cdots$$

which is one way to understand (30); hence there is a linear relationship between, for example, $h''(0)$ and $f''(0)$, where the proportionality factor is $\gamma^2$.

6.3. **Linearity In Differential Operators.** Along the same lines, the following observation is used in ODE's. If for any sufficiently differentiable function $f$ we define

$$(\mathcal{L}f)(x) \stackrel{\text{def}}{=} f''(x) + p(x)f'(x) + q(x)f$$

for any continuous functions $p, q$, then we easily check that $\mathcal{L}$ takes sufficiently differentiable functions to continuous functions and satisfies

$$\mathcal{L}(\alpha f + \beta g) = \alpha \mathcal{L}(f) + \beta \mathcal{L}(g)$$

for any $\alpha, \beta \in \mathbb{R}$ and $f, g$ sufficiently differentiable. This type of linearity is used in solving linear ODE's. Similarly, linearity is a crucial idea in solving linear PDE's.

## 7. Classical PageRank and Markov Chains

*Markov chains* is a large area of research. A simple illustration of their usefulness are the classical *PageRank* algorithms.

7.1. **Simplified PageRank.** Google used to publicly disclose the way they ranked webpages in order of "importance," which is one ingredient in how they decide which websites you see when you type in a search term (the other main ingredient is "relevance" of a webpage to the term you type). This created a bit of a "cat and mouse" game: Google would announce changes to its ranking algorithm (and its relevance algorithm) to provide a better search engine, whereupon website designers would change their websites to improve their chances of coming within the top ten pages that Google displays for various search terms. The first algorithm used by Google was called *PageRank*, possibly named after one of its two founders. Here is a simplified version.

**Example 7.1.** Imagine that the internet consists of four webpages, $A, B, C, D$, where

(1) page $A$ has one link to each of the pages $B, C, D$;
(2) page $B$ has one link to each of pages $A, C$;
(3) page $C$ has one link to each of pages $A, D$; and
(4) page $D$ has one link to each of pages $A, B$.

PageRank works as follows: we think of taking a "random walk" on this internet, where a walker who is on page $A$ takes a "random step" to one of $B, C, D$, each with "probability $1/3$"; similary for every webpage we imagine that a walker chooses one of the links on the page and randomly jumps there. This yields an array of numbers:

| | $A$ | $B$ | $C$ | $D$ |
|---|---|---|---|---|
| $A$ | 0 | $1/3$ | $1/3$ | $1/3$ |
| $B$ | $1/2$ | 0 | $1/2$ | 0 |
| $C$ | $1/2$ | 0 | 0 | $1/2$ |
| $D$ | $1/2$ | $1/2$ | 0 | 0 |

This gives a *Markov matrix* or probability matrix for the associated *Markov chain*

$$P = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix} .$$

In Math 223 we will learn how to multiply matrices; it will then be easy to see that $P^2 = PP$ represents what happens after we walk for two steps on this Markov chain, and $P^{100}$ represents what happens after 100 steps.

To see what happens as we walk for progressively more steps on this Markov chain, we can use Julia or MATLAB: we type

```
P = [0  1/3  1/3  1/3;  1/2  0  1/2  0;  1/2  0  0  1/2;  1/2  1/2  0  0]
```

and then issue the commands `P ^2` and `P^100` to examine $P^2$ and $P^{100}$ (for these very simple commands, same syntax works both in Julia and MATLAB). We see that for large $n$, such as $n = 100$, Julia reports $P^{100}$ as

```
44 Array\{Float64,2\}:
 0.333333  0.222222  0.222222  0.222222
 0.333333  0.222222  0.222222  0.222222
 0.333333  0.222222  0.222222  0.222222
 0.333333  0.222222  0.222222  0.222222
```

We will understand to mean that (1) the PageRank of webpage $A$ is $1/3$, and those of $B, C, D$ are $2/9$; (2) each row of $P^{100}$ represents the *stochastic vector* representing the unique *stationary distribution*.

**Exercise 7.1.** Following the previous example, Google's strategy is annouced or leaked. The designers of webpage $B$ learns of Google's strategy, whereupon they introduce 10 links from webpage $B$ to itself; now webpage $B$ has 12 links, 10 to itself, and 1 to each of $A$ and $C$. The new Markov matrix is

$$\texttt{P} = [\texttt{0 1/3 1/3 1/3; 1/12 10/12 1/12 0; 1/2 0 0 1/2; 1/2 1/2 0 0}]$$

What is $P^2$, $P^{100}$, and the new PageRank? (exercise).

**Exercise 7.2.** Google observes some webpage designers artificially boosting their PageRank and decides not to count any links from a webpage to itself. Google's revised strategy is annouced or leaked. The designers of webpage $B$ learns of Google's revised strategy, whereupon they collude with webpage $C$ as follows: $B$ adds 100 links to webpage $C$ (hidden at the bottom of their webpage), and $C$ add 100 links to webpage $B$. The new Markov matrix is

$$\begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/102 & 0 & 101/102 & 0 \\ 1/102 & 100/102 & 0 & 1/102 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix}.$$

What is $P^2$, $P^{100}$, $P^{1000}$ and the new PageRank (exercise)?

**Exercise 7.3.** Google observes webpage designers artificially boosting $B$ and $C$'s PageRank and decides not to count more than one link from any webpage to any other. Google's revised strategy is annouced or leaked. The designers of webpages $B, C$ learn of Google's revised strategy, whereupon they decide that webpage $B$ will only have one link to $C$ (and no other links), and vice versa for webpage $C$. The new Markov matrix is

$$\begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix}. \tag{31}$$

What is $P^2$, $P^{100}$, $P^{101}$, $P^{102}$, and $P^{103}$? Can you explain why you see what you see for $P^n$ and $n$ large (e.g., $n = 100, \ldots, 103$)?

**Exercise 7.4.** Google observes that some webpages have no way to reach other webpages. They decide to replace $P$ based on webpage links with

$$Q = (.85)P + (.15)E$$

where

$$\begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}.$$

What does $E$ represent? What is $Q^2$ and $Q^{100}$ for the $P$ of (31)?

**Exercise 7.5.** Google stops announcing and leaking its PageRank algorithm.

7.2. **Markov Matrices.** *Markov Chains* is a large field of study; part of this field involves *Markov matrices*, which is a generalization of the above PageRank examples. A Markov matrix is any matrix of the form

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

where $p_{ij}$ represents the probability of transitioning from state $i$ to state $j$; hence each row of $P$ is a *stochastic vector*, meaning a collection of non-negative numbers whose sum equals one. It is well-known that Markov chain that is *irreducible* (i.e., from any state you can reach any other in some number of steps) has a unique *stationary distribution*, meaning a stochastic vector $(\pi_1, \ldots, \pi_n)$ such that for each $i$ we have

$$\pi_i = \sum_{j=1}^{n} \pi_j p_{ji} \ . \tag{32}$$

**Exercise 7.6.** The children in your home are running in a circle from room A to room B to room C to room D and back to A (ad infinitum, or seemingly so). You model their trajectory by the Markov chain

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

What are $P^{10}, P^{11}, P^{12}, P^{13}$ ? What is the unique stationary distribution?

**Exercise 7.7.** Which Markov chain(s) in the above PageRank are not irreducible?

**Exercise 7.8.** Find stationary distributions of matrices $P$ above by typing in Julia: `using LinearAlgebra` and then `nullspace(transpose(P - I))` or `nullspace( (P-I)' )`; in MATLAB type `null( transpose( P - eye(4) ))` or `null( ( P - eye(4) )')`

An (irreducible) Markov chain is called *reversible* if $\pi_j p_{ji} = \pi_i p_{ij}$ for all $i, j$; in Math 223 we will be convinced that this implies that you could run the Markov chain "in reverse" and it would "look the same" running in reverse. Some physical situations are "time reversible" (e.g., $n$ celestial bodies moving according to Newton's law of gravitation) and some are not (e.g., a movie where someone increases entropy by spilling a glass of milk). (Add exercises.)

**Exercise 7.9.**

7.9(a) Show that $\vec{x} = \vec{x}(t)$ is a vector-valued function $\mathbb{R} \to \mathbb{R}^n$ of a single variable $t$ that satisfies the equations

$$\frac{d^2}{dt^2}\vec{x} = \vec{F}(\vec{x}),$$

for some vector-valued function $F\colon \mathbb{R}^n \to \mathbb{R}^n$, then $\vec{y}(t) \stackrel{\text{def}}{=} \vec{x}(-t)$ satisfies the same equation, i.e.,

$$\frac{d^2}{dt^2}\vec{y} = \vec{F}(\vec{y}).$$

7.9(b) Explain the relevance of the above observation to Newtonian celestial mechanics, where point masses experience a gravitation force that depends only on the position of the point masses. Are the equations of Newtonian mechanics "time reversible"?

## 8. Graphs, Constrained Data, and Regular Languages

8.1. **$(2,7)$-Constrained Data.** Some decades ago IBM introduced a strategy to prevent errors in magnetic storage: the idea was to first convert general binary strings (i.e., sequences of 0's and 1's) into longer $\{0,1\}$ strings that satisfied certain constraints that reduced the number of read errors. Errors in reading tended to occur when there were (1) too few 0's between any two consecutive 1's, and (2) too many 0's. Specifically, say that a binary string is $(2,7)$-*constrained* if between every two consecutive 1's there are between two and seven 0's. IBM converted each binary string of length $n$ into a string of length $2n$ that is $(2,7)$-constrained and stored the longer string. The reason for choosing $2,7$ in the constraint was that it (1) significantly reduced the chance of error, and (2) there are simple algorithms to convert an arbitrary binary string of length $n$ into a $(2,7)$-contrained string of length $2n$, and to convert the longer string back to the original string.

8.2. **More Details on the Motivation Behind $(2,7)$-Constrained Data.** A sequence of 0's and 1's were stored linearly on a magnetic tape, which one can view as a sequence of vertical magnets on a long horizontal strip, each magnet polarized either with $+$ on top and $-$ on bottom, or vice versa; the 1 was stored as an instruction to change the polarity, and a 0 as to keep the same polarity. When 1's were stored near one another, the quick changes in polarity tended to cause errors; when there were too many zeros between two 1's, the clocks that measured when a polarity change occurred were not accurate enough (e.g., to distinguish between seven 0's and eight zeros 0's).

8.3. **The Number of $(2,7)$-Constrained Words.** A matrix computation based on graph theory shows that the number of $(2,7)$-constrained strings of length $n$ is proportional $2^{\gamma n}$, where $\gamma > 0.5$. If $\gamma$ were strictly less than $1/2$, then it would be impossible to translate strings of length $n$ (of which there are $2^n$) into $(2,7)$-constrained strings of length $2n$, since the number of such strings is proportional to $2^{2n\gamma}$ which would be too small. Since $\gamma > 1/2$, for sufficiently large $n$ there are more than $2^n$ $(2,7)$-strings of length $2n$, and hence there is an injection—and hence an encoding and decoding—from the binary strings of length $n$ to $(2,7)$-strings of length $2n$. However there are a number of very efficient algorithms to carry this out; furthermore results from symbolic dynamics give similarly efficient algorithms for many generalizations of $(2,7)$-strings.

8.4. **Directed Graphs as Modeling $(2,7)$-Strings.** A *directed graph* is a quadruple $(V,E,t,h)$ where $V,E$ are sets (finite in cases of interest here) and $t,h$ are maps $E \to V$; we refer to $V$ as the *vertex set*, $E$ as the *edge set*, and $t,h$ as, respectively, the *tails map* and the *head map*. The graph in Figure 1 model $(2,7)$-strings in a sense that we now explain. This graph has vertex set $\{v_0,\ldots,v_7\}$ and a directed edges with heads and tails indicated in Figure 1; we have also put labels on the directed edges (for simplicity we have drawn the edges whose head is $v_0$ in a way that indicates their common label 1).

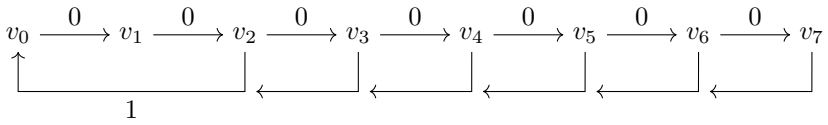


FIGURE 1. A directed graph (with labels on its edges) that models $(2,7)$-strings.

A *walk*, $k$, in a directed graph $G = (V, E, h, t)$ is an alternating sequence of vertices and edges

$$w = (v_0, e_1, v_1, e_2, v_2, \dots, e_k, v_k)$$

such that for all $i \in [k]$, $t(e_i) = v_{i-1}$ and $h(e_i) = v_i$; we refer to $k$ as the *length* of the walk, $w$. The labels on the edges in the directed graph in Figure 1 associate to each walk in this graph a $(2,7)$-string, and this association is surjective and takes at most six walks to any one string (exercise). The *adjacency matrix* of this graph,

$$A_G = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

has its $(i,j)$-th entry being the number of edges from the $i$-th vertex to the $j$-th vertex; equivalently, $A_G$ counts the number of walks of length one from one vertex to another, and in Math 223 we will see that $A_G^n$, the *$n$-th power of the matrix $A_G$*, tells us how many walks there are of length $n$ from any vertex of $G$ to any other (as a square array of numbers). In Math 223 we will state the *Perron-Frobenius theorem*, which will imply that the total number of walks of length $n$ in $G$ is *(asymptotically) proportional* to $\lambda_1^n$, where $\lambda_1 > 0$ is the *Perron-Frobenius eigenvalue* of $G$. It will follow that the number of $(2,7)$-strings of length $n$ is also propotional to $\lambda_1^n$.

**Exercise 8.1.** Type the above matrix into Julia or MATLAB; you can cut and paste from here, although likely you will need to add spaces (alternatively you can go to the course webpage):

```
A = [ 0  1  0  0  0  0  0  0 ;
0  0  1  0  0  0  0  0 ;
1  0  0  1  0  0  0  0 ;
1  0  0  0  1  0  0  0 ;
1  0  0  0  0  1  0  0 ;
1  0  0  0  0  0  1  0 ;
1  0  0  0  0  0  0  1 ;
1  0  0  0  0  0  0  0 ]
```

Compute $A^3$, and explain the entries that you see in terms of walks of length 3 in the graph in Figure 1. Set $B = A^{500}$, and compute a few entries of $B$ to the $1/500$ power. Are they larger than $\sqrt{2}$ or smaller? Why should we care? Type `eigs(A)` into MATLAB or `using LinearAlgebra` and `eigvals(A)` into Julia; do you see a number close to $\sqrt{2}$, and, if so, what is its value?

**Exercise 8.2.** Consider the "Fibonacci graph"


Write down its adjacency matrix, $A$. Compute the powers $A^2, A^3, A^4, A^5, A^6$. Have we seen this matrix before? Based on your observations, write down a formula for the number of walks of length $n$ from $v_0$ to itself, from $v_1$ to itself, from $v_0$ to $v_1$, and from $v_1$ to $v_0$. Prove by induction $n$ that your formulas are correct. Enter this matrix into Julia or MATLAB, and type `eigs(A)` into MATLAB or `using LinearAlgebra` and `eigvals(A)` into Julia; how close are these numbers to $(1 \pm \sqrt{5})/2$?

8.5. **Regular Languages.** An *alphabet* is a finite set; fix an alphabet $\mathcal{A}$ of size $a$. A *discrete finite automaton* (or simply *DFA*) *over* $\mathcal{A}$ is a digraph (directed graph) $G = (V, E, t, h)$ with the following additional structure, constraints, and conventions: (1) each vertex, $v$, has exactly $a$ edges leaving it (i.e., whose tails are $v$); (2) there is a "labelling" of the edges, which is a map $\ell\colon E \to \mathcal{A}$ such that each vertex has its leaving edges labelled with distinct elements of $\mathcal{A}$; (3) we also refer to the vertices of $G$ as the *states* of $G$; (4) there is a designed "starting state," $v_{\text{init}}$ of $V$, and (5) there is a designated subset of $V$ called the *final states* or *accepting states*. As such, there are exactly $a^k$ walksof length $k$ beginning at $v_{\text{init}}$, and the labelling associates to each walk a distinct *string* (or *word*) on $\mathcal{A}$ of length $k$ (i.e., a sequence of $k$ elements in $\mathcal{A}$). A string in $\mathcal{A}$ of length $k$ is *accepted* by this DFA if the associated walk beginning at $v_{\text{init}}$ ends in one of the final states, and is otherwise *rejected*; the subset of strings accepted by the DFA is the *language* (i.e., subset of all finite strings over $\mathcal{A}$) *recognized by the DFA*.

Etc.

## 9. Error Detection/Correction in Binary Data and ISBN Numbers

Error detection and correction is a large field of research. A basic example are a simple parity check for binary data and the check digit in ISBN numbers; this also illustrates the importance of finite fields and a fundamental usefulness of prime numbers (as opposed to composite numbers).

9.1. **Simple Parity Check.** One way to detect errors in the transmission of $n$-bits (binary digits, i.e., the digits $\{0, 1\}$) over a "noisy" channel is to add to the message $x_1, \ldots, x_n$ (with each $x_i = 0, 1$) a "parity check digit" $x_{n+1}$ given by the equation

$$x_1 + x_2 + \cdots + x_n + x_{n+1} \equiv 0 \pmod 2,$$

so that $x_{n+1} = 1$ if $x_1, \ldots, x_n$ contains an odd number of 1's, and otherwise $x_{n+1}$. In this way if a single $x_i$ is received incorrectly, then the resulting message has an odd number of 1's instead of an even number. In this way we can detect that there is has been an error in the message $x_1, \ldots, x_{n+1}$; we may be able to ask the transmitter to retransmit the message.

Of course, if there are two errors in the received message, or any even number of errors, then we cannot detect this.

9.2. **ISBN Numbers.** Until 2006, ISBN (International Stanard Book Number) Numbers would assign to each book a unique 9-digit identifier $x_1, \ldots, x_9$; the 10-th digit would be a "check digit" $x_10$ given by the equation

$$x_1 + 2x_2 + \ldots + 9x_9 + 10x_{10} \equiv 0 \pmod{11}$$

(if $x_{10}$ turned out to equal 10, this last digit was written with an $X$). This type of error detection allows us to detect if (1) a single digit is recorded/reported incorrectly, or (2) two (unequal) digits are interchanged (presumably the most common case is when the digits are consecutive) (exercise).

Modern ISBN numbers have 12 digits and a 13th check digit given by

$$x_1 + 3x_2 + x_3 + 3x_4 + \cdots + 3x_{12} + x_{13} \equiv 0 \pmod{10}$$

In this way we can detect an error if (1) a single digit is recorded/reported incorrectly, or (2) two (unequal) consecutive digits are interchanged, provided that they are not congruent modulo 5 (namely the pairs: $0, 5$, $1, 6$, ..., and $4, 9$)

9.3. **Error Correcting Codes.** In practice we often want to *correct* errors, not just detect them. For example, if we have a binary message $x_1, \ldots, x_n$ and we transmit a message that is three times as long:

$$x_1, x_1, x_1, x_2, x_2, x_2, x_3, \ldots, x_{n-1}, x_n, x_n, x_n$$

then if there is a single error in the reception of this message, we can not only detect it but recover the original message (by taking a majority vote for every group of three duplicates).

For $n = 1$, it is easy to see that if we want to send a single bit $x_1$ in a way that we can correct any single error in the received bits, we in fact need to send the message $x_1, x_1, x_1$ in place of $x_1$. However, if $n = 4$, rather than sending the message $x_1, x_2, x_3, x_4$ in 12 bits with each $x_i$ sent in triplicate, we claim that we can send the 7 bits

$$x_1, x_2, x_3, x_4, x_5, x_6, x_7,$$

where

$$\begin{aligned} x_5 &= x_1 + x_2 + x_3 \\ x_6 &= x_1 + x_3 + x_4 \\ x_7 &= x_1 + x_2 + x_4 \end{aligned}$$

and we can still correct any single error; to see this, we can verify (exercise, with hints) that each of the 16 possible values of $(x_1, x_2, x_3, x_4)$, the resulting 16 7-bit strings each are of *Hamming distance* at least 3 from each other, where for $\vec{x} = (x_1, \ldots, x_m)$ and $\vec{y} = (y_1, \ldots, y_m)$ we define

$$\rho_{\text{Hamm}}(\vec{x}, \vec{y}) \stackrel{\text{def}}{=} \big|\{i \in [m] \mid x_i \neq y_i\}\big|, \tag{33}$$

i.e., the Hamming distance of two strings is the number of components on which they differ.

The above way of transmitting 4 bits by adding 3 "check bits" is a famous *error-correcting code* known as a *Hamming code*. The code words are more easily

recognized as the *kernel* of the $3 \times 7$ matrix

$$H = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

where (1) $H$ is formed by writing all 7 nonzero strings of 3 bits, and (2) the *kernel* of $H$ refers to those strings of 7 bits, $\vec{z} = z_1, \ldots, z_7$ satisfying "$Hz = 0$" working over the integers modulo 2, i.e., the equations

$$\begin{aligned} 1z_1 + 0z_2 + 0z_3 + 1z_4 + 1z_5 + 0z_6 + 1z_7 &= 0 \\ 0z_1 + 1z_2 + 0z_3 + 1z_4 + 0z_5 + 1z_6 + 1z_7 &= 0 \\ 0z_1 + 0z_2 + 1z_3 + 0z_4 + 1z_5 + 1z_6 + 1z_7 &= 0 \end{aligned}$$

The fact that this kernel turns out to be 4-*dimensional* is why one can transmit 4 bits with this Hamming code.

**Exercise 9.1.** (1) Show that if $\vec{x} = (x_1, \ldots, x_m)$ and $\vec{y} = (y_1, \ldots, x_m)$ are two sequences of bits, then with $\rho_{\text{Hamm}}$ as in (33) we have

$$\rho_{\text{Hamm}}(\vec{x}, \vec{y}) = \rho_{\text{Hamm}}(\vec{0}, \vec{x} + \vec{y})$$

where $\vec{0} = (0, \ldots, 0)$ and

$$\vec{x} + \vec{y} = (x_1 + y_1, \ldots, x_m + y_m)$$

where the addition over bits is done modulo two (in logic this is the *exclusive-or* (XOR), where 1 is "true" and 0 is "false").

(2) Say that a set of binary strings of some length, $S \subset \{0,1\}^m$, is a *linear code* if for any $\vec{x}, \vec{y} \in S$ we have $\vec{x} + \vec{y} \in S$. Show that the Hamming code above is a linear code.

(3) Show that the Hamming code above has no code word of distance 1 or 2 to $\vec{0}$ (this will require a bit of case analysis).

(4) Show that the Hamming code above has *minimum distance* 3, i.e., any distinct strings of length 7 of the code have Hamming distance at least three.

(5) Conclude that if $\vec{x}$ is a code word and $\vec{y}$ is of distance at most one to $\vec{x}$, then there is no other codeword $\vec{x}'$ that is of distance at most one to $\vec{y}$.

**Exercise 9.2.** (1) Show that if $\vec{z} = (z_1, \ldots, z_7)$ is not a code word for the above Hamming code, i.e., if at least one of

$$\begin{aligned} 1z_1 + 0z_2 + 0z_3 + 1z_4 + 1z_5 + 0z_6 + 1z_7, \\ 0z_1 + 1z_2 + 0z_3 + 1z_4 + 0z_5 + 1z_6 + 1z_7, \\ 0z_1 + 0z_2 + 1z_3 + 0z_4 + 1z_5 + 1z_6 + 1z_7 \end{aligned}$$

is not zero, then there is a code word that is of distance 1 to $\vec{z}$.

(2) Since there are 8 strings $\{0,1\}^7$ that are of distance one or zero to any fixed element of $\{0,1\}^7$, and each element of $\{0,1\}^7$ is of distance zero or one to one of 16 words in the above code, is it possible that some element of $\{0,1\}$ is of distance one or zero to two distinct code words?

(3) Explain why the previous exercise implies that the above Hamming code is of minimum distance at least 3, and hence can correct at least 1 error in a received bit.

More generally, for any $r \in \mathbb{N}$ we can transmit $2^r - r - 1$ bits by adding $r$ "check bits" to get an error-correcting code each of whose codewords are of distance at least 3 from one another (and hence this code can correct a single error); we construct this using the analogous construction of a matrix $H$ with $2^r - 1$ columns that contain each nonzero string of length $r$. These *Hamming codes* are *perfect codes* such that the "Hamming distance balls of radius 1" cover all of the set of strings of $2^r - 1$ bits without (any intersection) and any leftover.

**Exercise 9.3.** Prove that the minimum distance of any Hamming code is at least three.

(Other exercises.)

## 10. Motivation from Graphics

## References

Department of Computer Science, University of British Columbia, Vancouver, BC V6T 1Z4, CANADA, and Department of Mathematics, University of British Columbia, Vancouver, BC V6T 1Z2, CANADA.

*E-mail address*: jf@cs.ubc.ca or jf@math.ubc.ca